Applications and Industry®

September 1961

Transactions Papers

NUMBER 56

Published Bimonthly by

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

Communication and Electronics—September 1961

Note to Librarians. The six bimonthly issues of "Applications and Industry," March 1961–January 1962, will also be available in a single volume (no. 80) entitled "AIEE Transactions—Part II. Applications and Industry," which includes all technical papers on that subject presented during 1961. Bibliographic references to Applications and Industry and to Part II of the Transactions are therefore equivalent.

Applications and Industry. Published bimonthly by the American Institute of Electrical Engineers, from 20th and Northampton Streets, Easton, Pa. AIEE Headquarters: 345 East 47th Street, New York 17, N. Y. Address changes must be received at AIEE Headquarters by the first of the month to be effective with the succeeding issue. Copies undelivered because of incorrect address cannot be replaced without charge. Editorial and Advertising offices: 345 East 47th Street, New York 17, N. Y. Nonmember subscription $8.00 per year (plus 75 cents extra for foreign postage payable in advance in New York exchange). Member subscriptions: one subscription at $5.00 per year to any one of three divisional publications: Communication and Electronics, Applications and Industry, or Power Apparatus and Systems: additional annual subscriptions $8.00 each. Single copies when available $1.50 each. Second-class mail privileges authorized at Easton, Pa. This publication is authorized to be mailed at the special rates of postage prescribed by Section 132.122.

Printed in United States of America

Number of copies of this issue 5,100

Problems of Asymptotic Behavior and Stability

LAMBERTO CESARI
NONMEMBER AIEE

THIS PAPER attempts to focus on a number of areas in nonlinear differential equations which appear to be of heightened interest today.

When a physical phenomenon is reduced to a mathematical one, a certain degree of schematization and simplification necessarily occurs, and countless details, considered as inessential, are disregarded. Here those phenomena are of interest where the inherent nonlinearity cannot be disregarded. But, when it is assumed that a system of differential equations

$$\dot{x}=f(x, t) \tag{1}$$

$x=(x_1, \ldots, x_n)$, $f=(f_1, \ldots, f_n)$, $(\dot{})=d/dt$, "represents" a given phenomenon, it must not be overlooked that at least the numerical coefficients appearing in f are known only with a certain degree of approximation, and that even the form of f is not certain. Also, if a certain solution $X(t)$ is defined by its initial values $X(0)$, it must be considered that $X(0)$ may never be known exactly. In addition, to know $X(0)$ may not be of interest. Finally, the physical system is constantly under the influence of countless disturbances which cannot be taken into consideration. All this leads to the conclusion that if a solution $X(t)$ of equation (1) is to have any physical interest, it must possess two kinds of stability:

A stability with respect to the initial values so that all solutions $\tilde{X}(t)$ corresponding to initial values $\tilde{X}(0)$ sufficiently "close" to $X(0)$ are in some sense close to $X(t)$ for $t, 0\leq t<+\infty$. There are a number of different mathematical formulations of this requirement: Lyapunov stability, asymptotic stability, orbital stability, asymptotic orbital stability, etc. The problem in itself suggests which of these concepts has to be applied.

2. A stability of the differential system itself [in particular with respect to the parameters and physical constants involved, and at least in a neighborhood of the solution $X(t)$], so that the solutions $x=y(t)$ of any system $\dot{x}=g(x,t)$ sufficiently close to $\dot{x}=f(x,t)$, are close to and present corresponding behavior of the solutions $x=x(t)$ of $\dot{x}=f(x,t)$. The only mathematical formulation so far known of this more difficult requirement is the structural stability of Andronov and Pontryagin.[1] (For information on the general subject, see references 1, 2, and 3.)

Realistic as all this seems to be, it must not be overlooked that all kinds of regulators, control systems, and servomechanisms present some small time lag, and even some degree of heredity in their actual working. Therefore, the mathematical problem should actually be a more complicated one. These further difficulties must be disregarded here. Here, primarily nonlinear differential systems of the form

$$\dot{x}=f(x, t) \tag{1}$$

are considered or, in the autonomous case, of the form

$$\dot{x}=f(x) \tag{2}$$

It is well known how futile it is to try to solve the system formally or numerically. In most situations, however, this is not needed, but only the determination of the steady states of the system (points of equilibrium, periodic solutions, family of periodic solutions, invariant surfaces, etc.) and, more precisely, only those which are stable. The difficulties connected with the actual solutions of nonlinear differential systems impose the use of qualitative methods.

Points of Equilibrium and Their Stability in the Small

First an autonomous differential system in considered:

$$\dot{x}=f(x), \text{ i. e., } \dot{x}_i=f_i(x_1, \ldots, x_n), \quad i=1, \ldots, n, \tag{3}$$

where $x=(x_1, \ldots, x_n)$, $f=(f_1, \ldots, f_n)$. Let $||x-y||$ denote as usual the Euclidean distance of two points x and y, so that $||x||$ is the distance of the point x from the origin. Here a point x is of course the representative of the state of a system in the phase space. A point $x_0=(x_{01}, \ldots, x_{0n})$ where

$$f(x_0)=0, \text{ i.e.,}$$

$$f_i(x_{01}, \ldots, x_{0n})=0, \quad i=1, \ldots, n, \tag{4}$$

is said to be a point of equilibrium of the system. The points of equilibrium correspond to the constant solutions of the system, e.g., $x=x_0$, i.e., $x_i=x_{0i}$, $i=1, \ldots, n$, for all $-\infty<t<+\infty$. The determination of the points of equilibrium does not involve the solution of the differential system of equations 3, but only the numerical solution of the finite system of equations 4. Graphical and numerical devices for this purpose are well known. Once an isolated point of equilibrium x_0 is known, it is important to establish whether it is stable. This verification also, may not require the solution of the differential system. Let us consider the linear part in the development of $f(x)$ in Taylor series, around x_0, i.e., $f(x)=A(x-x_0)+R(x)$, since $f(x_0)=0$, and we have $R(x)/||x-x_0||\to 0$ as $x\to x_0$, and $A=(a_{ih})$ is an $n\times n$ constant matrix, $a_{ih}=\partial f_i/\partial x_h$ at $x=x_0$. An important theorem by Lyapunov states:

If the characteristic roots of the constant matrix A have negative real parts, then the point of equilibrium x_0 is asymptotically stable and the system is structurally stable in a neighborhood of x_0.

Under these conditions all solutions $x=x(t)$ of the system of equations 3, with initial point $x(0)$ sufficiently close to x_0, approach x_0 as $t\to+\infty$. The requirement concerning the matrix A means that the linear system with constant coefficients $\dot{u}=Au$ has all solutions approaching zero as $t\to+\infty$. The requirement on A itself is algebraic in character and can be verified by well-known devices.[4] In actual analysis it may be convenient to reduce x_0 to the origin by means of a displacement in the phase space.

For a nonautonomous system

$$\dot{x}=f(x, t) \tag{5}$$

presenting a constant solution, $x=0$, and hence $f(0, t)=0$ for all t, the relation holds $f(x)=A(t)x+R(x, t)$, and it is assumed that $R(x, t)/||x||\to 0$ uniformly, in t as $x\to 0$. Lyapunov's theorem holds also in this case, provided we replace the words "characteristic roots of A''' by type numbers of the solutions of the linear system (with variable coefficients) $\dot{u}=A(t)u$, and provided this is sufficiently regular. In the periodic case, the

Paper 60-1020, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Pacific General Meeting, San Diego, Calif., August 8–12, 1960, and presented for discussion only at the AIEE Winter General Meeting, New York, N. Y., January 29–February 3, 1961. Manuscript submitted October 31, 1959; made available for printing December 29, 1960.

LAMBERTO CESARI is with the University of Michigan, Ann Arbor, Mich.

This paper was written while the author was with the Research Institute for Advanced Studies, Baltimore, Md., under contract no. AF-49(638)-

type numbers are the usual characteristic exponents and the system is regular. Thus, even in the nonautonomous case, the analysis of the underlying linear system suffices for the asymptotic stability of the constant solution $x=0$ of the differential system 5.

If one characteristic root of the constant matrix A is zero, or two characteristic roots are purely imaginary, stability and instability may depend upon the nonlinear terms of the development of $f(x)$. Conclusive theorems of Lyapunov may answer the question of the stability of the point of equilibrium.[5,6,7]

For $n=2$, for instance, for mechanical systems with one degree of freedom, much more information is available than for $n>2$.[3,7,8]

The Second Method of Lyapunov and Stability in the Large

The theorem mentioned previously concerns only stability in the small, i.e., whether a system slightly deviated from a position of equilibrium x_0 will return toward x_0 at $t\to+\infty$. None of these theorems give any indication as to how small the deviations have to be in order that the system return toward x_0. Stability in the large is a more difficult problem. Obviously, a necessary condition for stability in the large is that there is only one point of equilibrium x_0 in the whole phase space. Both questions may be answered by using the second method of Lyapunov. The underlying idea consists of finding a function $V=V(x)=V(x_1,\ldots,x_n)$ in a neighborhood U of the point of equilibrium (which may be assumed to be $x=0$), or in the whole phase space, satisfying the following two conditions: $V(0)=0$ and $V(x)$ is definitely positive, i.e., $V(x)>0$ for $x\neq 0$; $V'(x)=\Sigma_i f_i(x)\partial V/\partial x_i<0$ for all $x\neq 0$. If a function V can be found satisfying these two conditions, then the point $x=0$ is asymptotically stable.[5]

If the conditions for V are satisfied in the whole phase space and either $V\to+\infty$ as $||x||\to+\infty$, or $V'\leq -m<0$ for all $||x||\geq M$, and some m, $M>0$, then the point $x=0$ is asymptotically stable in the large. Otherwise, the underlying analysis may determine a neighborhood U_0, in general smaller than U, such that the stability is assured for all possible deviations within U_0. Practical consideration may determine whether the size of U_0 is sufficient. It may be helpful to know that V' is actually the total derivative of V with respect to t along the solutions of the system. Hence, $V'<0$ states that V is a decreasing function of t.

The theorem mentioned is the prototype of a series of analogous theorems by Lyapunov and others, concerning both autonomous and time-dependent systems.[5-7,9,10]

For instance, the system

$$\dot{x}_1=x_2-x_1x_2^2-x_1^3,\ \dot{x}_2=-x_1-x_1^2x_2-x_2^3$$

presents only one point of equilibrium $x_1=x_2=0$. Letting $V=x_1^2+x_2^2$ yields $V>0$ and $V'=-2(x_1^2+x_2^2)^2<0$ for all $(x_1,x_2)\neq(0,0)$. Thus the origin is asymptotically stable in the large.

By means of the second method of Lyapunov, A. I. Lure[10] has studied a wide class of regulated systems with one regulating organ. For the sake of simplicity such a regulator is considered here, represented by a differential system of the form

$$\dot{x}_s=-r_sx_s+f(y),\ s=1,\ldots,n$$
$$\dot{y}=b_1x_1+\ldots+b_nx_n-f(y) \qquad (6)$$

where y represents the regulator and $f(y)$ a nonlinear function characteristic of the regulator, with $f(0)=0$, $yf(y)>0$ for $y\neq 0$. Thus $x_1=\ldots=x_n=y=0$ is a position of equilibrium. If all constants r_s, b_s are supposed to be positive, then the condition $\Sigma_s b_s/r_s<1$ assures the asymptotic stability of the position of equilibrium; see reference 6, where $f(y)$ is supposed to be of the form $f(y)=gy^m+g_1y^{m+1}+\ldots$, m odd, $g>0$.

Assume ϵ is a positive small constant and consider the differential equation[6]

$$\dot{x}=\epsilon^2x-x^3$$

The points of equilibrium are 0, $+\epsilon$, $-\epsilon$, and it can easily be seen that 0 is unstable while $+\epsilon$ and $-\epsilon$ are stable. A moving point displaced from the origin will remain within the interval $[-\epsilon,\ +\epsilon]$ and, if pushed outside this interval, will return toward $+\epsilon$ or $-\epsilon$ at $t\to+\infty$. Since ϵ is supposed to be small, the origin may be regarded as "stable" for all practical purposes.

Consider instead the differential equation:[6]

$$\dot{x}=-\epsilon^2x+x^3$$

Again, 0, $+\epsilon$, $-\epsilon$ are the points of equilibrium, 0 is stable, $\pm\epsilon$ are unstable, and a moving point, pushed outside the small interval $[-\epsilon,\ +\epsilon]$ will approach either $+\infty$, or $-\infty$ as $t\to+\infty$. Thus the origin, although theoretically stable, may be regarded as "unstable" (for all practical purposes).

Even in such questions the consideration of a V function may be of interest. Assume for instance, that a function V can be found satisfying both conditions, $V>o$, $V'<o$, in the whole phase space except a small neighborhood U of the origin, and with $V\to+\infty$ as $||x||\to+\infty$. Then there will be another neighborhood U_0 of the origin, containing U and which is assumed to be small too, such that the moving point x having departed from U_0 will return toward U_0 as $t\to+\infty$. In other words, the origin is "stable" and even "stable in the large" if small (although unpredictable) oscillations within a small neighborhood U_0 of the origin are disregarded. For instance, if a, ϵ are positive constants, and ϵ small, consider the system

$$\dot{x}_1=-x_2+a\epsilon x_1-ax_1^3$$

$$\dot{x}_2=x_1+a\epsilon x_2-ax_2^3$$

Letting $V=x_1^2+x_2^2$ yields $V'=-2a(x_1^4+x_2^4)+2a\epsilon(x_1^2+x_2^2)$. Outside the circle with center at the origin and radius $2\sqrt{\epsilon}$ there is $\sqrt{2\epsilon}<\max(|x_1|,\ |x_2|)$, and hence

$$-V'/2a=(x_1^4+x_2^4)-\epsilon(|x_1|^2+|x_2|^2)\geq$$
$$(x_1^4+x_2^4)-2\epsilon\max(|x_1|^2,\ |x_2|^2$$
$$>(x_1^4+x_2^4)-\max(|x_1|^4,\ |x_2|^4)\geq$$
$$(x_1^4+x_2^4)-(x_1^4+x_2^4)=$$

Then, outside U, $V>0$, $V'<0$, and V, $x_1^2+x_2^2$ must decrease. This implies that a representative point x displaced from the circle U will return toward $U=U_0$ as $t\to+\infty$. Thus the origin may be regarded as "stable," even in the large, small oscillations of amplitude less than $4\sqrt{\epsilon}$ around the origin are disregarded. (A more detailed analysis would show that the origin is unstable and there is a stable cycle of small diameter.) For further discussion on the general subject of Lyapunov's second method see references 6 and 7.

Periodic Solutions

1. It may occur that an autonomous system $\dot{x}=f(x)$, $x=(x_1,\ldots,x_n)$, has some isolated periodic solution (cycle), of some period T, or a family of periodic solutions (cycles) whose period depends upon the amplitude; or that a system $\dot{x}=f(x,t)$ with f periodic in t of a given period T has periodic solutions whose dominating terms are of period T (harmonics), or of period Tm (subharmonics), or of period T/m (ultraharmonics), (m is an integer). More complicated situations also may occur. The existence of these periodic solutions together with further important information may be obtained without solving the system.

For $n=2$ these existence theorems are very general and apply as well to the situations usually labelled as relaxation

cillations. They are usually given for the second-order Liénard equation. For the autonomous case the following quite general theorem of N. Levinson is pertinent (see reference 2, p. 174).

Consider the second-order autonomous differential equation

$$\ddot{x}+f(x,\dot{x})\dot{x}+g(x)=0 \qquad (7)$$

i.e., the system $\dot{x}=v$, $\dot{v}=-f(x,v)v-g(x)$, where f and g are continuous functions of their arguments. Suppose that there are positive constants a, m, M such that $f(x,v)\geq m$ if both $|x|\geq a$, $|v|\geq a$, that $f(x,v)\geq -M$ for all x and v and $f(0,0)=0$. Suppose that $xg(x)>0$ for all $x\neq 0$, and $g(x)\to\pm\infty$, $g(x)/G(x)\to 0$ as $x\to\pm\infty$, where $G(x)=\int_0^x g(u)du$. Then there exists at least one nonconstant periodic solution of equation 7.

For instance, consider the van der Pol equation $\ddot{x}+\mu(x^2-1)\dot{x}+x=0$, where μ is any positive constant. This equation regulates the oscillations of a rather typical feedback mechanism (e.g., a feedback circuit with triode as exemplified in reference 2, p. 125). Here $g(x)\equiv x$, and hence $xg(x)=x^2>0$ for all $x\neq 0$, $g(x)\to\pm\infty$ as $x\to\pm\infty$; $G(x)=x^2/2$, and hence $g(x)/G(x)=1/2x\to 0$ as $x\to\pm\infty$; $f(x,v)=\mu(x^2-1)\geq -\mu$ for all x and v, and $f(x,v)\geq\mu>0$ for all $|x|\geq 2$ and all v. Thus, the van der Pol equation satisfies the conditions of the theorem and the existence of a cycle is assured, no matter how large the positive constant μ. [By other analogous considerations (e.g. reference 2, p. 143), this cycle is shown to be unique and asymptotically orbitally stable.]

For the Liénard equation with a periodic forcing term $e(t)$ (see reference 2, p. 176):

$$\ddot{x}+f(x,\dot{x})\dot{x}+g(x)=e(t) \qquad (8)$$

the same theorem holds for the existence of harmonics. In both situations further statements are known concerning the uniqueness of the cycle or of the periodic solution, and evaluations of the amplitude and of the period (in the autonomous case). In the situation described in the theorems mentioned, the periodic solution or cycle already is certainly stable if it is unique. A great many theorems are known concerning the Liénard equations and 8.[2,3,8]

2. For $n\geq 2$ a great deal of information is available for systems containing a small parameter, say either of the form

$$\dot{x}=Ax+\epsilon f(x,\epsilon) \text{ (autonomous case)} \qquad (9)$$

or of the form

$$\dot{x}=Ax+\epsilon f(x,t,\epsilon) \qquad (10)$$

where ϵ is the small parameter, A a constant $n\times n$ matrix, and $f(x,t,\epsilon)$ periodic in t of some period $T=2\pi/\omega$, $x=(x_1,\ldots,x_n)$, $f=(f_1,\ldots,f_n)$.

For instance, if A is an $n\times n$ constant matrix whose characteristic roots have negative real parts, if $f(x,t)$ is continuous in its arguments, periodic in t of period T, with bounded incremental ratios with respect to $x_1,\ldots,x_n$, and $f(0,t)=0$, and if $g(t)$ is periodic in t of period T, then the system

$$\dot{x}=Ax+xf(x,t)+\epsilon g(t) \qquad (11)$$

has an asymptotically stable periodic solution of period T (harmonic). (For this theorem, due to A. B. Farnell, C. E. Langenhop, and N. Levison, see, e.g., reference 2, p. 94.) The reader may also refer to references 2, 3, and 8 for literature on the subject. The well-known methods of B. van der Pol, N. M. Krylov, and N. Bogolyubov will not be discussed here.

3. The author and J. K. Hale[2, 11–17] have developed an approach which has led both to existence and stability theorems for periodic solutions and to a practical process of successive approximations for their determination. Here is a short examination of this process, and following it some of the existence theorems and applications to well-known differential problems. For more details on the process of successive approximations and applications, see the expositions in reference 2 or the papers quoted previously. The method of successive approximations is better described in relation to a differential system of the form

$$\dot{y}=Ay+\epsilon q(y,t,\epsilon),\ y=(y_1,\ldots y_n),\quad q=(q_1,\ldots,q_n) \qquad (12)$$

where ϵ is a small parameter, $|\epsilon|\leq\epsilon_0$, A an $n\times n$ matrix whose elements are constants, or continuous functions of ϵ, and q an $n\times n$ matrix whose elements are continuous functions of $y_1,\ldots,y_n$ with bounded incremental ratios, and periodic functions of t of period $T=2\pi/\omega$, integrable in $[0,T]$ (or alternatively independent of t and then system 12 is autonomous). For the sake of brevity assume that A is a diagonal matrix $A=\operatorname{diag}(\rho_1,\ldots,\rho_n)$ where the numbers ρ_j may depend on ϵ, are continuous functions of ϵ, and $\rho_j(0)=i\tau_j=ia_j\omega/b_j$, a_j, b_j integers, $b_j>0$, for $j=1,\ldots,m$, $1\leq m\leq n$, and, if $m<n$, also $\rho_j(0)\neq ih\omega/b$, $b=b_1b_2\ldots b_m$, $i=\sqrt{-1}$, for all $j=m+1,\ldots,n$, and $h=0,\pm 1,\pm 2,\ldots$. One may try to find a solution "close" to a solution of the linear system $z'=A(0)z$, of the form

$$z(t)=(c_1e^{i\tau_1 t},\ldots,c_me^{-i\tau_m t},0,\ldots,0)$$

$c_j\neq 0$ constants. Let $B=\operatorname{diag}(i\tau_1,\ldots,i\tau_m,\rho_{m+1},\ldots,\rho_n)$, let $\phi(t)=(\phi_1,\ldots,\phi_n)$ be any vector function of period $2\pi b/\omega$, and $m[e^{i\tau_j t}\phi_j(t)]=c_j$, $j=1,\ldots,m$, where m denotes the usual mean value. Let $\psi=F\phi$ be the transformation defined by

$$\psi(t)=F[\phi(t)]=z(t)+\epsilon e^{Bt}\int e^{-Bu}\times\{q[\phi(u),u,\epsilon]-D[\phi]\phi(u)\}du \qquad (13)$$

where $D=\operatorname{diag}(d_1,\ldots,d_m,0,\ldots,0)$ is the diagonal $n\times n$ matrix defined by

$$c_jd_j=m\{e^{i\tau_j t}q_j[\phi(t),t,\epsilon]\},\ j=1,\ldots,m$$

Here e^{Bt} is the $n\times n$ diagonal matrix whose diagonal elements are $\exp(i\tau_j t)$, $j=1,\ldots,m$, $\exp\rho_j t$, $j=m+1,\ldots,n$, while e^{-Bt} is the inverse of e^{Bt} and can be obtained by changing t into $-t$. In equation 13 ψ is a vector function $\psi=(\psi_1,\ldots,\psi_n)$. For each component ψ_j, $j=1,\ldots,m$, the integrand in equation 13 is periodic of mean value zero and the integral denotes the unique primitive of mean value zero; for each component ψ_j, $j=m+1,\ldots,n$, the integrand is of the form $\exp(-\rho_j t)M(t)$, $M(t)$ periodic, and the integral is the unique primitive of the same form For $|\epsilon|$ sufficiently small, the method of successive approximations

$$y^{(0)}(t)=z(t),\ y^{(k+1)}(t)=F[y^{(k)}(t)],\quad k=0,1,2,\ldots \qquad (14)$$

converges toward a vector function $y(t)$ which satisfies the functional equation $y=F[y]$, i.e., $y(t)$ is the fixed element of the transformation $\psi=F[\phi]$. Also, $y(t)$ satisfies the differential system

$$\dot{y}(t)=(B-\epsilon D)y+\epsilon q(y,t,\epsilon)$$

with $D=D[y]$. Thus $y(t)$ is a periodic solution of the given system 12 provided the relations hold (determining system)

$$B-\epsilon D=A,\text{ or } ia_j\omega/b_j-\epsilon d_j(a,b,c,\omega,\epsilon)=\rho_j,\ j=1,\ldots,m \qquad (15)$$

The analysis of the functional transformation $\psi=F[\phi]$, or of the method of successive approximations, defined by equations 14 and of the determining equations 15, has given direct proofs of existence theorems, some of which are mentioned, with examples, subsequently.

4. The following will discuss one of the existence theorems. Consider the real autonomous differential system

$$\begin{aligned}&\ddot{x}_1+\sigma_1{}^2x_1=f_1(x,\dot{x},\sigma)\\&\ddot{x}_j+2\alpha_j\dot{x}_j+\sigma_j{}^2x_j=\sigma f_j(x,\dot{x},\epsilon),\ j=2,\ldots,m\\&\dot{x}_j+\beta_jx_j=\epsilon f_j(x,\dot{x},\epsilon),\ j=m+1,\ldots,n\end{aligned} \qquad (16)$$

where $2\leq m\leq n$, $x=(x_1,\ldots,x_n)$, $\dot{x}=(x_1,\ldots,x_m)$, $f_1,\ldots,f_n$ are continuous functions of their arguments with bounded incremental ratios, $\sigma_j(\epsilon)>0$, $\alpha_j(\epsilon)\geq 0$, $\beta_j(\epsilon)>0$ are continuous functions of ϵ,

and we put $\sigma_1(0)=\omega_0$. Assume that for each $j=2, \ldots, m$ either $\alpha_j(0)>0$, or $\alpha_j(0)=0$, then $k\omega_0\neq\sigma_j(0)$, $k=0, 1, \ldots$. In the existence theorem stated subsequently a function of the real variable $\lambda>0$ is needed which is denoted by $P(\lambda, \omega_0, 0)$ and given by

$$P(\lambda,\omega,0)=(T\lambda)^{-1}\int_0^T f_1[\lambda\omega^{-1}\sin\omega t,0,\ldots, 0,\lambda\cos\omega t,0,\ldots,0,0]\cos\omega t dt \quad (17)$$

for every ω, and $T=2\pi/\omega$. The theorem states:

If the equation $P(\lambda,\omega_0,0)=0$ has a simple root for $\lambda=\lambda_0>0$, then the differential system of equation 12 has a cycle (periodic solution) of the form

$$x_1(t,\epsilon)=\lambda(\epsilon)\omega^{-1}(\epsilon)\sin\omega(\epsilon)t+O(\epsilon)$$
$$x_j(t,\epsilon)=O(\epsilon),\ j=2,\ldots,n \quad (18)$$

for all ϵ sufficiently small and where $\lambda(\epsilon)$, $\omega(\epsilon)$ are continuous functions of ϵ with $\lambda(0)=\lambda_0$, $\omega(0)=\omega_0$.[15] (See also reference 2, p. 128, and reference 11.)

Under the same conditions of the aforementioned theorem the functions $\lambda(\epsilon)$, $\omega(\epsilon)$ can be obtained by the numerical solution of a system of finite equations (the determining equations) of the form

$$P(\lambda,\omega,\epsilon)=0,$$
$$\omega-\epsilon Q(\lambda,\omega,\epsilon)=\sigma_1(\epsilon),$$

where the functions $P(\lambda, \omega, \epsilon)$, $Q(\lambda, \omega, \epsilon)$ can be obtained by the method of successive approximations which, at the same time, give approximations of the periodic solution. Here $P(\lambda, \omega, 0)$ is given by equation 17, and $Q(\lambda, \omega, 0)$ by

$$Q(\lambda,\omega,0)=(T\lambda)^{-1}\int_0^T f_1[\lambda\omega^{-1}\sin\omega t,0,\ldots, 0,\lambda\cos\omega t,0,\ldots,0,0]\sin\omega t dt \quad (19)$$

Under the hypotheses of the same theorem the condition for orbital stability of the periodic solution 18 of system 16 is very simple: The cycle 18 is orbitally stable for $t\to+\infty$ provided

$$\int_0^T f_{j\dot{x}_j}dt<0 \quad (20)$$

for $j=1$, and for each $j=2, \ldots, m$, with $\alpha_j(0)=0$.[14] Here $f_{j\dot{x}_j}$ means the partial derivative of f_j with respect to $\dot{x}_j$ where the arguments of $f_{j\dot{x}_j}$ are taken along the periodic solution 18 for $\epsilon=0$.

Note that all that is needed in the existence theorem is $P(\lambda, \omega, 0)$ which is given by equation 17 by means of a quadrature. Furthermore, the computation of $P(\lambda, \omega, 0)$ may be simplified by the following remarks. First, letting $Z(x_1, \dot{x}_1)=f(x_1, 0, \ldots, 0, \dot{x}_1, 0, \ldots, 0, 0)$ and decomposing Z into its even and odd components, e.g., $Z=Z_1+Z_2+Z_3+Z_4$, with Z_1 even in x_1 and odd in $\dot{x}_1$, Z_2 even in x_1 and $\dot{x}_1$, Z_3 odd in x_1 and $\dot{x}_1$, Z_4 odd in x_1 and even in $\dot{x}_1$, then $P(\lambda, \omega, 0)$ depends only on Z_1. If Z_1 is a polynomial

$$Z_1=\Sigma a_{hk}x_1^{2h}\dot{x}_1^{2k-1},\ h\geq 0,\ k\geq 1 \quad (21)$$

then[11]

$$P(\lambda,\omega_0,0)=2^{-2}\Sigma a_{hk}\frac{(2h)!(2k)!}{h!k!(h+k)!}\left(\frac{\lambda}{2\sigma_1}\right)^{2h+2k-2} \quad (22)$$

For instance, consider again the van der Pol equation $\ddot{x}+\epsilon(x^2-1)\dot{x}+x=0$, where now ϵ is a positive parameter sufficiently small. Applying the theorems mentioned previously the system 12 is reduced to its first equation, $\ddot{x}+x=\epsilon(1-x^2)\dot{x}$, hence, $1=m=n$, $\sigma_1\equiv 1$, $\omega_0=1$, $f_1=(1-x^2)\dot{x}$, and the remaining hypotheses are automatically satisfied since there are no other equations. Since f_1 is even in x and odd in $\dot{x}$, $f_1=Z=Z_1$, and Z_1 is a polynomial as in equation 21 with $a_{01}=0$, $a_{11}=-1$, $a_{hk}=0$ otherwise. Hence from equation 22 we have $P(\lambda, \omega_0, 0)=(1/2)(1-\lambda^2/4)$, and the equation $P(\lambda, \omega_0, 0)=0$ yields $\lambda=2$. Thus, according to the existence theorem, the van der Pol equation has a cycle

$$x(t,\epsilon)=\lambda\omega^{-1}\sin(\omega t+\gamma)+O(\epsilon),\ \gamma \text{ arbitrary},$$
$$\lambda(\epsilon)=2+O(\epsilon),\ \omega(\epsilon)=1+O(\epsilon) \quad (23)$$

i.e., with amplitude close to 2 and frequency close to 1. Also, $f_{1\dot{x}}=1-x^2$, the integral of condition 20 becomes

$$\int_0^{2\pi}[1-(2\sin t)^2]dt=-2\pi<0$$

and thus asymptotic orbital stability is assured.

The expressions 23 could have been obtained also very easily by the first step of the method of successive approximations defined by equations 14 (see reference 2, p. 130). The second step of the same method gives (see ref. 2, p. 130):

$$x(t,\epsilon)=\lambda\omega^{-1}(\sin(\omega t+\nu)-(\epsilon/32)\lambda^3\omega^{-1}\times\cos 3(\omega t+\nu)+O(\epsilon^3)$$

As a further example, consider the system of the van der Pol type of differential equations:

$$\ddot{x}_1+x_1=\epsilon(1-x_1^2-x_2^2)\dot{x}_1+\epsilon f_1(x_1,x_2,\dot{x}_2)+\epsilon g_1(x_1,\dot{x}_2,x_2,\dot{x}_2)x_2$$
$$\ddot{x}_2+2x_2=\epsilon(1-x_1^2-x_2^2)\dot{x}_2+\epsilon f_2(x_1,\dot{x}_1,x_2)+\epsilon g_2(x_1,\dot{x}_1,x_2,\dot{x}_2)x_1 \quad (24)$$

where to avoid computations, $f_1(-x_1, x_2, \dot{x}_2)=-f_1(x_1, x_2, \dot{x}_2)$, $f_2(x_1, \dot{x}_1, -x_2)=-f_2(x_1, \dot{x}_1, x_2)$ is assumed. Applying the theorem yields $2=m=n$, $\sigma_1=1$, $\sigma_2=2^{1/2}$, $\omega_0=1$, and hence $k\omega_0=k\neq 2^{1/2}=\sigma_2$ for all $k=0, 1, \ldots$. The integral of equation 17 where x_1 is replaced by $\lambda\omega^{-1}\sin\omega t$, $\dot{x}_1$ by $\lambda\cos\omega t$, and $x_2, \dot{x}_2$ by zero, becomes the sum of three parts, of which the one concerning g_1 is zero since $x_2\equiv 0$, the one concerning f_1 is zero since the integrand is odd in t, and thus

$$P(\lambda,\omega,0)=(T\lambda)^{-1}\int_0^T(1-\lambda^2\omega^{-2}\sin^2\omega t)\lambda\times\cos^2\omega t dt=(1/2)(1-\lambda^2/4\omega^2$$

where $T=2\pi/\omega$. For $\omega=\omega_0=1$, $P=(1/2)(1-\lambda^2/4)$ and $P=0$ yields $\lambda=2$. Hence the system 24 has the periodic solution (cycle)

$$(1)\ x_1=\lambda\sin(\omega t+\nu)+O(\epsilon),\ x_2=O(\epsilon),\ \omega=1+O(\epsilon),\ \lambda=2+O(\epsilon$$

By exchanging the office of the two equations 24 we have $\sigma_1=2^{1/2}$, $\sigma_2=1$, $\omega_0=2^{1/2}$, and hence $k\omega_0\equiv k2^{1/2}\neq 1=\sigma_1$ for all $k=0, 1, \ldots$. By replacing x_2 by $\lambda\omega^{-1}\sin\omega t$, $\dot{x}_2$ by $\lambda\cos\omega t$, x_1 and $\dot{x}_1$ by zero the same expression for P results and hence, for $\omega=\omega=2^{1/2}$, $P=(1/2)(1-\lambda^2/8$ $\lambda=2^{3/2}$. It is concluded that the system 24 has also the periodic solution

$$(2)\ x_1=O(\epsilon),\ x_2=\lambda\sin(\omega t+\gamma),\ \omega=2^{1/2}+O(\epsilon),\ \lambda=2^{3/2}+O(\epsilon$$

By applying equation 20 it could easily be seen that both cycles, 1 and 2, are asymptotically orbitally stable for $t\to+\infty$.

One of the existence theorems for periodic solutions of systems of type will be mentioned below. It refers to real systems of the type

$$\ddot{x}_j+\sigma_j^2x_j=\epsilon f_j(x,\dot{x},t,\epsilon),\ j=1,\ldots,m,$$
$$\ddot{x}_j+2\alpha_j\dot{x}_j+\sigma_j^2x_j=\sigma f_j(x,\dot{x},t,\sigma),\ j=m+1,\ldots,n \quad (2$$

where $1\leq m\leq n$, $x=(x_1, \ldots, x_n)$, $\dot{x}=(\ldots, \dot{x}_n)$, the functions f_j are continuous in their space arguments and ϵ, periodic of period $T=2\pi/\omega$ in t, and integrable in $[0, T]$. Assume that $\alpha_j(\epsilon)\geq 0$, $\sigma_j(\epsilon)$ continuously differentiable functions of ϵ, or constants, $\alpha_j<\sigma_j$, and that certain integers a_j, b_j yield $\sigma_j(0)=a_j\omega/b_j$, $j=\ldots, m$, and either $\alpha_j(0)>0$, or $\alpha_j(0)=\sigma_j(0)\neq k\omega/b_0$, $b_0=b_1b_2\ldots b_m$, $k=0,1,\ldots$, $j=m+1, \ldots, n$. Certain functions Q_j, $j=1, \ldots, m$, of $\lambda=(\lambda_1, \ldots, \lambda_m)$, $\theta=(\theta_1, \ldots, \theta_m)$, ω and ϵ, will be needed which can be obtained by a process of successive approximations. Nevertheless, in the existence theorem which follows, only the function $P_j(\lambda, \theta, \omega, 0)$ are needed which are given by the integrals

$$P_j(\lambda,\theta,\omega,0)=(\lambda_jTb_0^{-1})[\cos\theta_j\int_0^{Tb_0}f_j\times\cos a_jb_j^{-1}\omega u du-\sin\theta_j\int_0^{Tb_0}f_j\sin a_jb_j^{-1}\omega u$$

$$Q_j(\lambda,\theta,\omega,0)=(\lambda_jTb_0^{-1})[-\sin\theta_j\int_0^{Tb_0}f_j\times\cos a_jb_j^{-1}u du-\cos\theta_j\int_0^{Tb_0}f_j\times\sin a_jb_j^{-1}\omega u du]$$

For the sake of simplicity it is assumed that the functions P_j, Q_j have continuous first partial derivatives, though much less is needed. The theorem states:

the system of finite equations

$$P_j(\lambda,\theta,\omega_0,0)=0,\ Q_j(\lambda,\theta,\omega_0,0)=-\sigma_j'(0),\quad j=1,\ldots,m$$

s a solution $\lambda=\lambda_0$, $\theta=\theta_0$ and the Jacobian the P_j, Q_j with respect to λ, θ at $\lambda=\lambda_0$, $=\theta_0$ is not zero, then the system 25 has a riodic solution of the form

$$x_j(t,\epsilon)=\lambda_j a_j^{-1}b_j\omega\sin(a_jb_j^{-1}\omega t+\theta_j)+O(\epsilon),\quad j=1,\ldots,m$$

$$x_j(t,\epsilon)=O(\epsilon),\ j=m+1,\ldots,n \tag{27}$$

r all ϵ sufficiently small, where $\lambda(\epsilon)$, $\theta(\epsilon)$ e continuous functions of ϵ with $\lambda(0)=\lambda_0$, $)=\theta_0$.[2,11,13]

he functions $\lambda(\epsilon)$, $\theta(\epsilon)$ are given by the termining equations

$$P_j(\lambda,\theta,\omega,\epsilon)=0$$

$$b_j^{-1}\omega-\epsilon Q_j(\lambda,\theta,\omega,\epsilon)=\sigma_j(\epsilon),\ j=1,\ldots,m \tag{28}$$

Note that for ϵ small the periods $T_j=b_j/a_j\omega$ of the dominant terms of the mponents $x_j(t,\epsilon)$ of the periodic solution are close, but not necessarily equal, the periods $2\pi/\sigma_j(\epsilon)$ of the free oscillaons of the harmonic oscillators $\ddot{x}_j+\sigma_j^2(\epsilon)x_j=0$, $j=1,\ \ldots\ m$. Thus the termining equations 28 state that the riod of the oscillations of the whole stem 25 is "locked" with the one of the ternal forces. This is the well-known enomenon of entrainment of frequency, pical in nonlinear resonance.

In the theorem just mentioned the case th all $a_j=b_j=1$ corresponds to harnic oscillations, and the case with $a_j>1$, =1 to subharmonic oscillations. eorems of stability analogous to the es mentioned have been recently proved J. K. Hale.[14]

By means of the described theorems d using the same general method, Hale, mbill, and Bailey[13,14,17] have disssed in detail, the harmonic, subrmonic, and ultraharmonic solutions existence and stability of a number of uations and systems of equations as, instance, the following ones:

$$\ddot{x}+\sigma^2x=B\cos 2\omega t+\epsilon\alpha\cos 2\omega t\,.\,x+\epsilon bx^3$$

$$\ddot{x}+x=\epsilon(1-x^{2n})\dot{x}+\epsilon\mu\omega\cos(\omega t+\alpha)$$

$$\ddot{x}+\sigma^2x=\epsilon[ax+bx\cos 2\omega t+cx^3+dx^2\dot{x}+ex^3\cos 2\omega t]$$

$$\ddot{x}_1+\sigma_1^2x_1=\epsilon(\alpha-\beta x_2^2)\dot{x}_1+p\cos t,$$

$$\ddot{x}_2+\sigma_2^2x_2=\epsilon(\nu-\delta x_1^2)\dot{x}_2+q\cos 2t$$

r further theorems and other applicans of the same method see references nd 11–17.

milies of Periodic Solutions

By the same method developed by author and J. K. Hale mentioned previously, it was possible to prove the existence of large families of periodic solutions (or cycles) for differential systems presenting convenient symmetries. Consider, for instance, an autonomous differential system of the form

$$\ddot{x}_j+\sigma_j^2x_j=\epsilon f_j(x,\dot{x},\epsilon),\ j=1,\ldots,\mu,$$

$$x_j=\epsilon f_j(x,\dot{x},\epsilon),\ j=\mu+1,\ldots,n \tag{29}$$

where ϵ is a small parameter, $x=(x_1,\ \ldots,\ x_n)$, $\dot{x}=(\dot{x}_1,\ \ldots,\ \dot{x}_n)$, and the functions f_j are continuous with bounded incremental ratios with respect to each of their arguments. Assume that $f_1(0, x_2,\ldots x_n, 0, \dot{x}_2, \ldots, \dot{x}_\mu, \epsilon)=0$, and that all f are odd in the vector $(x_1, \ldots, x_\mu)$. Let $\sigma_j(\epsilon)>0$, $m\sigma_j(0)\neq\omega_0$ for all $j=2, \ldots, \mu$, and $m=0, 1, \ldots$, where $\omega_0=\sigma_1(0)$. The theorem states:[11,12,16]

Under these conditions, for all ϵ sufficiently small, the system of equations 29 has a family of periodic solutions, depending upon $n-\mu+2$ arbitrary parameters λ_1, $\eta_1,,\ldots,$ $\eta_{n-\mu}$, γ, of the form

$$x_1(t,\epsilon)=\lambda_1\omega^{-1}\cos(\omega t+\gamma)+O(\epsilon)$$

$$x_j(t,\epsilon)=O(\epsilon),\ j=2,\ldots,\mu$$

$$x_j(t,\epsilon)=\eta_{j-\mu}+O(\epsilon),\ j=\mu+1,\ldots,n \tag{30}$$

where the period $2\pi/\omega$ is a continuous function of ϵ, $\lambda_1,\eta_1,\ldots$, $\eta_{n-\mu}$, and $\omega(0,\lambda_1,\eta_1,\ldots,\eta_{n-\mu})=\omega_0$.

An analogous statement holds if f_1 is even and f_j, $j=2, \ldots, \mu$, are odd in $(x_2, \ldots, x_\mu, x_1)$ where $\cos(\omega t+\gamma)$ is replaced by $\sin(\omega t+\gamma)$ in equation 30.

Consider, for instance, the equation $\ddot{x}+x=\epsilon|\dot{x}|x$, with $\epsilon>0$ small. Here $1=\mu=n$, $f_1=f_1(x,\dot{x})=|\dot{x}|x$, and f_1 is odd in x, and $f(0, 0)=0$. Thus this equation has a family of periodic solutions of the form $x=\lambda\omega^{-1}\cos(\omega t+\gamma)+O(\epsilon)$, $\omega=\omega(\lambda,\epsilon)$, $\omega(\lambda, 0)=1$, λ, γ arbitrary. (Note that f_1 is also even in $\dot{x}$, and hence the same periodic solutions can be written also in the form $x=\lambda\omega^{-1}\sin(\omega t+\gamma')+O(\epsilon)$, λ, γ' arbitrary.)

As a further example consider the system $\ddot{x}+x=\epsilon(1-|y|)x$, $\ddot{y}+2y=\epsilon(1-|\dot{x}|)y$, with $\epsilon>0$ small. Here $2=\mu=n$, $\sigma_1\equiv 1$, $\sigma_2\equiv 2^{1/2}$, $m\sigma_2=m2^{1/2}\neq 1\equiv\sigma_1$ for all m. Also, $f(x, y, \dot{x}, \dot{y})=(1-|y|)x$, $g(x, y, \dot{x}, \dot{y})=(1-|\dot{x}|)y$, both f and g are odd in (x, y) and $f(0, y, 0, \dot{y})=0$. Thus this system has a family of periodic solutions of the form $x=\lambda\omega^{-1}\cos(\omega t+\gamma)+O(\epsilon)$, $y=O(\epsilon)$, $\omega=\omega(\lambda,\epsilon)$, $\omega(\lambda, 0)=1$, λ, γ arbitrary, of frequencies close to 1. By exchanging the office of the two equations, we have $\sigma_1=2^{1/2}$, $\sigma_2=1$, $m\sigma_2\equiv m\neq 2^{1/2}=\sigma_1$ for all $m=0, 1, \ldots$, both f and g are odd in (x, y), and $g(x, 0, \dot{x}, 0)=0$. Thus the same system has another family of periodic solutions of the form $x=O(\epsilon)$, $y=\lambda\omega^{-1}\cos(\omega t+\gamma)+O(\epsilon)$, $\omega=\omega(\lambda,\epsilon)$, $\omega(\lambda,0)=2^{1/2}$, λ, γ arbitrary, of frequencies close to $2^{1/2}$.

The last-mentioned theorem, and others,[11,16] show the possibility of manifolds of periodic solutions not much dissimilar to those which are well known in the linear case.

Almost Periodic Solutions

It is known that if A is an $n\times n$ constant matrix whose characteristic roots have all negative real parts and $f(t)$ is an almost periodic function (input) then the linear differential system

$$\dot{x}+Ax=f(t)$$

has one and only one solution (output) which is almost periodic and asymptotically stable. This is true to some degree for nonlinear systems containing a small parameter. For instance, if $f(x, t, \epsilon)$, $g(x, t, \epsilon)$ are continuous vector functions in x with bounded incremental ratio with respect to each space variable, and they are almost periodic in t, with $f(0, t, \epsilon)=0$ $g(x, t, 0)=0$, then the nonlinear differential system

$$\dot{x}+Ax=f(x,t,\epsilon)+g(x,t,\epsilon) \tag{31}$$

has a stable almost periodic solution.[2,11]

The case where the matrix A has some zero or purely imaginary characteristic roots is more difficult and it will not be discussed here. Instead a type of research will be mentioned which has had a remarkable development in the last 10 years. It concerns the interaction of periodic phenomena of different periods, generally in irrational ratio.

Assume that a physical system S is regulated by a nonlinear differential system $\dot{x}=F(x)$ which has a stable periodic solution (cycle) $x=\phi(t)$ of some period T and thus S has free oscillations of that period. Assume that the system is perturbed and that the perturbed system S' is regulated by a differential system

$$\dot{x}=F(x)+\epsilon G(x,t) \tag{32}$$

where G is a smooth function of its arguments, and is periodic of some period $T'\neq T$. A theorem of Levinson[18] states:

If $x=\phi(t)$ is a strongly stable solution for the unperturbed system, then the perturbed system (equation 32) has, for all ϵ sufficiently small, a stable manifold (torus) of solutions.

Under these conditions, the manifold is covered by solutions which, in general, are not periodic but almost periodic, and each one passes as close as is desired to each point of the manifold. In the phase space x we may see these solutions as "close" to the unperturbed periodic cycle $x=\phi(t)$ as we want for ϵ sufficiently small. Each of the new solutions, com-

pared with the other solutions on the manifold, is quite unstable. On the other hand, the manifold as a whole is stable, and a representative point, not on the manifold, will return toward it as $t \to +\infty$. The structure may be considered "stable" for all practical purposes. Strong stability in Levinson's theorem means that all but one of the characteristic exponents of the usual linear variational system for $x = \phi(t)$ have negative real parts. For instance the equation, $\ddot{x}+3(x^2-1)\dot{x}+x=\epsilon \sin\sqrt{2}t$, for $|\epsilon|$ sufficiently small has a torus of almost periodic solutions, close to the stable cycle of the van der Pol equation $\ddot{x}+3(x^2-1)\dot{x}+x=0$. (Compare with section "Periodic Solutions" of this paper.)

Levinson's result has been widely extended by Russian authors, particularly Bogolyubov and Mitropolskii,[19] who have taken into consideration the case of $G(x, t)$ almost periodic in t, or presenting an arbitrary behavior. An improved formulation of their result is contained in a recent paper by Hale.[20] These theorems show the possibility of manifolds of solutions (invariant manifolds). These manifolds are of practical interest only if, in some sense, they are stable. The inherent relevant concept of stability is not easy to state and there are several degrees of stability which may be of actual interest. Research has been done on this subject and very simple criteria for stability have been given recently by Hale and Stokes.[21]

References

1. Singular Points of Differential Equations (in Russian), **A. Andronov, L. S. Pontryagin.** *Proceedings*, Akademii Nauk SSSR, Moscow, USSR, vol. 14, no. 2, 1937, pp. 247–50.

2. Asymptotic Behavior and Stability Problems in Ordinary Differential Equations (book), **L. Cesari.** Julius Springer, Berlin, Germany, 1959.

3. Theory of Ordinary Differential Equations (book), **Earl A. Coddington, Norman Levinson.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

4. Applied Analysis (book), **C. Lanczos.** Prentice-Hall, Inc., Englewood Cliffs, N. J., 1956.

5. Problème Général de la Stabilité du Mouvement, **A. Lyapunov.** *Communications*, Societé Mathematique de Krakov, Krakow, Poland, vol. 3, 1893, pp. 265–72; also *Annals of Mathematics Studies*, Princeton, N. J., vol. 17, 1949.

6. Theory of Stability of Motion (book, translated from Russian), **I. G. Malkin.** Office of Technical Services, Dept. of Commerce, Washington, D. C., 1952.

7. Differential Equations; Geometric Theory (book), **S. Lefschetz.** Interscience Publishers, Inc., New York, N. Y., 1957.

8. Equazioni Differenziali Nonlineari (book), **G. Sansone, R. Conti.** Roma Cremonese, Rome, Italy, 1956.

9. Stability of Nonlinear Regulatory Systems (book, in Russian), **A. M. Letov.** Gosteizdat, Moscow, USSR, 1955.

10. Some Nonlinear Problems of the Theory of Automatic Regulation (book, in Russian), **A. I. Lure.** Izdat, Moscow-Leningrad, USSR, 1951.

11. Existence Theorems for Periodic Solutions of Nonlinear Lipschitzian Differential Systems and Fixed Point Theorems, **L. Cesari.** "Contributions to the Theory of Nonlinear Oscillations," Princeton, N. J., vol. 5, 1960, pp. 115–172.

12. A New Sufficient Condition for Periodic Solutions of Weakly Nonlinear Differential Systems, **L. Cesari, J. K. Hale.** *Proceedings*, American Mathematical Society, Providence, R. I., vol. 8, 1957, pp. 757–64.

13. Subharmonic and Ultraharmonic Solutions for Weakly Nonlinear Systems, **R. A. Gambill, J. K. Hale.** *Journal of Rational Mechanics and Analysis*, Bloomington, Ind., vol. 1956, pp. 353–98.

14. On the Stability of Periodic Solutions of Weakly Nonlinear Periodic and Autonomous Differential Systems, **J. K. Hale.** "Contributions to the Theory of Nonlinear Oscillations," Princeton, N. J., vol. 5, 1960, pp. 91–114.

15. Periodic Solutions of Nonlinear Systems of Differential Equations, **J. K. Hale.** *Rivista di Matematica*, Parma, Italy, vol. 5, 1954, pp. 281–311.

16. Sufficient Conditions for the Existence of Periodic Solutions of Systems of Nonlinear First and Second Order Differential Equations, **J. K. Hale.** *Journal of Mathematics and Mechanics*, Bloomington, Ind., vol. 7, 1958, pp. 163–72.

17. On the Stability of Periodic Solutions of Weakly Nonlinear Differential Equations, **H. R. Bailey, R. A. Gambill.** *Journal of Rational Mechanics and Analysis*, vol. 6, 1957, pp 655–68.

18. Small Periodic Perturbations of Autonomous System with a Stable Orbit, **Levinson.** *Annals of Mathematics*, Princeton, N. J., vol. 52, 1950, pp. 727–28.

19. Asymptotic Methods in the Theory of Nonlinear Oscillations (book, in Russian), **N. Bogolyubov, Yu. A. Mitropolsky.** Gosteizdat, Moscow, USSR, second edition, 1958.

20. Integral Manifolds of Perturbed Differential Systems, **J. K. Hale.** *Annals of Mathematics*, Princeton, N. J., vol. 73, no. 3, 1961.

21. Behavior of Solutions Near Integral Manifolds, **J. K. Hale, A. Stokes.** *Archive for Rational Mechanical Analysis*, Berlin, Germany, vol. 6, 1960, pp. 133–70.

22. Stability Theory of Differential Equations (book), **Richard Bellman.** McGraw-Hill Book Company, Inc., New York, N. Y., 1953.

23. Control System Analysis and Design via the "Second Method" of Lyapunov. I. Continuous Time Systems; II. Discrete-Time Systems, **J. E. Bertram, R. E. Kalman.** *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 82, series D, 1960, pp. 371–93, 394–400.

24. Some Problems in the Theory of Nonlinear Oscillations, Vols. I and II (book, translated from Russian), **I. G. Malkin.** Office of Technical Services, Dept. of Commerce, Washington, D. C., 1956.

Optimum Control of Nonlinear Discrete-Data Systems

JULIUS T. TOU
MEMBER AIEE

BOONYOK VADHANAPHUTI
NONMEMBER AIEE

Synopsis: This paper introduces a technique for the synthesis of nonlinear discrete-data control systems to fulfill an optimal performance criterion. The synthesis procedure is developed by use of the state-transition method. Digital compensation is designed for the nonlinear system to have deadbeat performance. The technique presented permits the use of a digital computer to carry out the design of optimum control.

DURING the past decade, many techniques have been developed for the analysis and design of discrete-data control systems.[1–3] Practically all the design methods are derived on the basis of linearity. Digital control systems of many types and with very real practical interest tend increasingly to require the use of nonlinear equations in their mathematical description, in place of the much simpler linear equations which have often sufficed in the past. Nonlinear sampled-data control systems have been analyzed by several approaches.[4–8] Known techniques for analyzing such nonlinear systems are limited essentially to three: the describing-function method, the phase-plane method, and numerical methods. The describing-function method, when applied to a system containing a nonlinear element and a sampler, is generally of limited usefulness. The method of phase-plane analysis has more pertinent application but is limited, practically to plants with a second-order transfer function. The numerical methods would involve a fair number of computations, although the complexity of the plant imposes no limitation.

Because of the difficulties with nonlinear problems, little work has been done on the synthesis of nonlinear digital or sampled-data control systems. Mu[ller] has presented an interesting paper[9] on compensation of saturating sampl[ers]

Paper 61-76, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 29–February 3, 1961. Manuscript submitted August 23, 1960; made available for printing December 2, 1960.

Julius T. Tou and Boonyok Vadhanaphuti are with Purdue University, Lafayette, Ind.

The work reported was supported in part by the Office of Naval Research under contract *Nonr 1100(18)*, through the Information Systems Branch.

ta control systems. His design procedure is straightforward but lengthy. It volves a certain amount of guesswork, d would face some difficulties when plied to complex systems with other nlinearities than idealized saturation or niting. This paper introduces a simpler d more systematic method for the synesis of nonlinear discrete-data feedback ntrol systems to meet an optimal perrmance criterion. The development the synthesis procedure makes use of atrices and the state-transition techque.[10-12] Digital compensation is dened for the nonlinear system to have adbeat performance. This technique ay also be used to design nonlinear conol for the compensation of linear and nlinear systems.

ate-Transition Analysis

Analysis and synthesis of linear conol systems may, generally, be carried t from two major approaches. One, e commonly used block diagram apoach, involves the determination of the nsfer characteristics of the system mponents and the over-all transfer aracteristic. The other is based upon e characterization of a system by a mber of simple first-order differential uations describing the state variables, th the initial conditions given by the te-transition equations. This is ially carried out with a state-variable gram and may be referred to as the te diagram approach.

While the block diagram approach is erally limited to the design of linear tems, the state diagram approach may extended to the synthesis of certain es of nonlinear systems. To familiarinterested readers with this method of alysis and the terminology to be used in er development, a brief review of the te-transition technique is presented follows.

A linear system can be described by a of first-order linear differential equans, which may be expressed in vector m as

$$=\|A\|\mathbf{v}(\lambda) \tag{1}$$

ere $\lambda=t-nT$ and $0\leq\lambda\leq T$. Equan 1 is often referred to as the state ferential equation of the system; a column matrix for the state variables nd y

$$\left\|\begin{matrix}\mathbf{x}\\ \mathbf{y}\end{matrix}\right\| \tag{2}$$

d is referred to as the state vector. The ut and the system state variables are denoted by x and y respectively. The initial conditions for the state differential equation may be given in vector form as

$$\mathbf{v}(nT^+)=\|B\|\mathbf{v}(nT) \tag{3}$$

This equation, which describes the transition of the system state variables at the sampling instants, is referred to as the state-transition equation. Both $\|A\|$ and $\|B\|$ are square matrices and can be written down by inspection of the state diagram of the system. The state diagram is essentially the same as the analog computer simulation diagram for the system, which is made up of integrators, summing amplifiers, potentiometers, samplers, clamps, and simple delay elements.[1]

Taking the Laplace transform of equation 1 gives

$$s\mathbf{V}(s)=\|A\|\mathbf{V}(s)+\mathbf{v}(0^+) \tag{4}$$

Rearranging yields

$$\mathbf{v}(s)=\|sI-A\|^{-1}\mathbf{v}(0^+) \tag{5}$$

Taking the inverse transform of equation 5 yields the solution to the state differential equation as

$$\mathbf{v}(\lambda)=\|\phi\|\mathbf{v}(0^+) \tag{6}$$

where $\|\phi\|$, defined as the transition matrix of the system, is given by

$$\|\phi\|=e^{A\lambda}=L^{-1}(\|sI-A\|^{-1}) \tag{7}$$

Equation 7 provides a formal way of determining the transition matrix. However, it can be found by a short cut. The elements of the transition matrix may be determined readily by inspection of the state diagram of the system.

In terms of t, equation 6 becomes

$$\mathbf{v}(t)=\|\phi(t-nT)\|\mathbf{v}(nT^+) \tag{8}$$

Combining with equation 3 yields

$$\mathbf{v}(t)=\|\phi(t-nT)\|\,\|B\|\mathbf{v}(nT) \tag{9}$$

which describes the system behavior during the interval $nT\leq t\leq(n+1)T$. Thus, at $t=(n+1)T$

$$\mathbf{v}(n+1T)=\|\phi(T)\|\,\|B\|\mathbf{v}(nT) \tag{10}$$

This is a recursion equation from which the values of the state variables at the sampling instants can be computed. Equations 9 and 10 provide the complete solution for the system performance.

Development of the Method

Consider the nonlinear sampled-data control system shown in Fig. 1(A). $G(s)$ is the transfer function of the linear part of the plant; N is the transfer characteristic of the nonlinear element of the plant; and

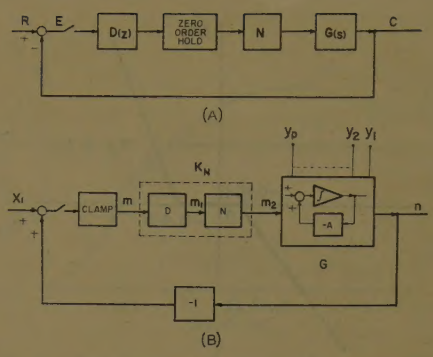


Fig. 1. Representations of discrete-data control system

A—Block diagram
B—State diagram

$D(z)$ is the digital controller to be determined. A zero-order hold is used as a data-smoothing device. This nonlinear system is designed for performing with deadbeat response to a step-function input.

The block diagram is redrawn as shown in Fig. 1(B), with the position of D and the clamp interchanged and the $G(s)$ described by its state-diagram representation. Let the signal output of the clamp be m, which is the clamped-error signal. The state vector of the system is given by

$$\mathbf{v}=\left\|\begin{matrix}x_i\\ \mathbf{y}\\ m\end{matrix}\right\| \tag{11}$$

in which

$$\mathbf{y}=\left\|\begin{matrix}u_1\\ y_2\\ \vdots\\ y_p\end{matrix}\right\|$$

is the vector for the state variables of the plant, $y_1, y_2, \ldots y_p$. The plant is assumed to be of the pth order and is described by p state variables. The symbol y_1 also represents the output of the control system. The input to system y_1 is assumed to be a step function.

Variable Gain Concept

The required digital controller and the system nonlinear element may be treated as a unit of variable gain K_N, and will be called the D-N combination. The variable gain K_N will have different values during different sampling periods; it is also dependent upon the characteristic of the nonlinear element N. The input to the D-N combination is m and the output is m_2. At any sampling instant $t=nT^+$, and m_2 and m are related by a constant K_n; thus

$$m_2(nT^+)=K_nm(nT^+) \tag{12}$$

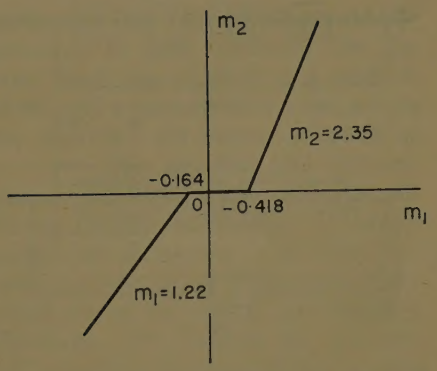


Fig. 2. Asymmetrical nonlinearity

K_n is the gain constant of the D-N combination during the $(n+1)$th sampling period.

Based upon the above argument, the transition matrix $\|\phi\|$ of the system is expressed as a function of the variable gain K_n, and will have different values at different sampling instants. From equations 3 and 10 it is obtained that at $n=0$

$$\mathbf{v}(0^+)=\|B\|\mathbf{v}(0) \tag{13}$$

$$\mathbf{v}(T)=\|\phi_0\|\mathbf{v}(0^+) \tag{14}$$

Since both $\|B\|$ and $\mathbf{v}(0)$ are known, $\mathbf{v}(0^+)$ is defined, and

$$m(0^+)=x_1(0^+) \tag{15}$$

The transition matrix $\|\phi_0\|$ is a function of K_0. If m (0^+) exceeds the saturation limit of the nonlinear element, K_0 should be equal to the maximum allowable value of the variable gain K_N during the first sampling period. Thus

$$K_0=\frac{m_2(0^+)}{x_1(0^+)} \tag{16}$$

where m_2 (0^+) is given by the saturation limit of the nonlinear element. At $n=1$

$$\mathbf{v}(T^+)=\|B\|\mathbf{v}(T) \tag{17}$$

from which the value of m (T^+) can be computed.

$$\mathbf{v}(2T)=\|\phi_1\|\mathbf{v}(T^+) \tag{18}$$

where $\|\phi_1\|$ is a function of K_1. If $m(T^+)$ is greater than the saturation limit, K_1 should be equal to the maximum allowable value of K_N during the second sampling period. K_1 is then given by the ratio between the saturation limit and $m(T^+)$.

At $n=2$

$$\mathbf{v}(2T^+)=\|B\|\mathbf{v}(2T) \tag{19}$$

$$\mathbf{v}(3T)=\|\phi_2\|\mathbf{v}(2T^+) \tag{20}$$

where $\|\phi_2\|$ is a function of K_2. If m $(2T^+)$ exceeds the saturation limit, K_2 may be determined in similar fashion. However, if $m(2T^+)$ is less than the saturation limit, the determination of K_2 calls for other conditions for deadbeat performance. It can readily be shown that the system error will be zero for $t\geq kT$, if

$$y_1(kT)=x_1(kT) \tag{21}$$

and

$$y_2(kT)=y_3(kT)=\ldots y_p(kT)=0 \tag{22}$$

where $y_1(kT)$, $y_2(kT)$, ... $y_p(kT)$ are functions of the successive constants of the variable gain K_N, and can be derived from

$$\mathbf{v}(kT)=\|\phi_{k-1}\|\mathbf{v}(\overline{k-1}T^+) \tag{23}$$

Hence, the successive constants for K_N can be obtained by solving equations 21 and 22.

Now, the input and output signals of the D-N combination are completely determined. The z-transforms of these two signals are

$$M(z)=m(0^+)+m(T^+)z^{-1}+\ldots+m(kT^+)z^{-k} \tag{24}$$

$$M_2(z)=K_0m(0^+)+K_1m(T^+)z^{-1}+\ldots+K_km(kT^+)z^{-k} \tag{25}$$

The z-transform of the input signal to the nonlinear element follows from equation 25:

$$M_1(z)=a_0+a_1z^{-1}+\ldots+a_kz^{-k} \tag{26}$$

Since a_k and $K_km(kT^+)$ are the input and output of the nonlinear element, the coefficients a_k's can be determined either analytically or graphically. Therefore, the required digital controller for deadbeat response is given by

$$D(z)=\frac{\sum_{j=0}^{k}a_jz^{-j}}{\sum_{j=0}^{k}m(jT^+)z^{-j}} \tag{27}$$

The above development of the synthesis technique is best illustrated by numeri examples, which are given in the follo ing section. From the discussion it noted that the settling time of the c trol system depends upon the order the plant as well as the nonlinear ch acteristic. The presence of saturation limiting may cause a longer settling ti

Synthesis Procedure

Based upon the technique just veloped, the synthesis of nonlinear crete-data control systems can be car out systematically in four major steps

1. Draw the state diagram of the syst with the desired digital controller and nonlinear element represented by a vari gain K_N, and determine the transi matrix and the B-matrix of the system inspection of the state diagram.

2. Evaluate the input signal to the combination from the transition matrix the B-matrix by using equations 3 and and determine the successive values of variable gain K_N to satisfy the requ ments for deadbeat performance.

3. Compute the output signal of the combination from the results of step 2, evaluate the input signal to the nonli element, analytically or graphically.

4. Determine the pulse-transfer func $D(z)$ by use of equation 27.

This synthesis procedure can also applied to the design of digital contro for multirate, variable-rate, and fi pulse-width sampled-data control syst containing nonlinear elements.

Illustrative Examples

To illustrate the synthesis proced several numerical examples are prese in the following.

Example 1, Asymmetrical Nonlinearity

Consider the nonlinear sampled- control system given in example reference 5. The transfer function $G(s)$ $1/s$ $(s+1)$. The nonlinear charac istic is reproduced as shown in Fig The sampling period is one second, the system is subjected to a unit- input. Design a digital controller this system to exhibit deadbeat respo Zero initial conditions are assumed.

The state diagram of this syste drawn in Fig. 3. Inspection of the s diagram yields the transition matrix

$$\|\phi(\lambda)\|=\begin{vmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 1-e^{-\lambda} & K_n(\lambda-1+e^{-\lambda})\\ 0 & 0 & e^{-\lambda} & K_n(1-e^{-\lambda})\\ 0 & 0 & 0 & 1\end{vmatrix}$$

and

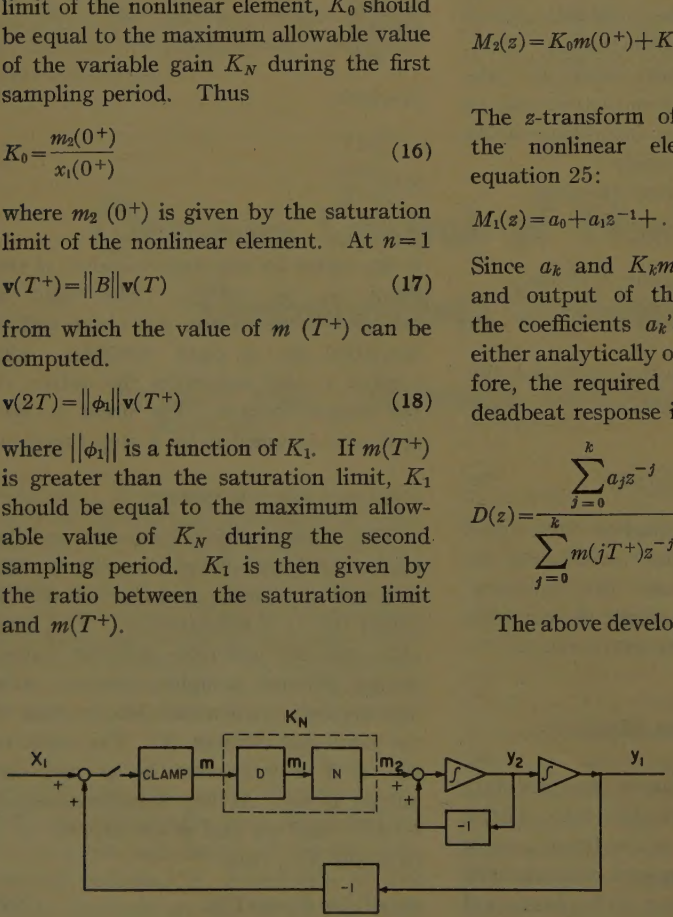


Fig. 3. State diagram of the illustrative example

$$T)\| = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0.632 & 0.368K_n \\ 0 & 0 & 0.368 & 0.632K_n \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

e B-matrix is found to be

$$= \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \end{Vmatrix}$$

en, for $n=0$

$$^+)=\|B\|\mathbf{v}(0)=\|1 \quad 0 \quad 0 \quad 1\|'$$

ch gives $m(0^+)=1$. The symbol $\| \|'$ gnates the transpose matrix. Since

$$= \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0.632 & 0.368K_0 \\ 0 & 0 & 0.368 & 0.632K_0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

$$)=\|\phi_0\|\mathbf{v}(0^+)=\|1 \quad 0.368K_0 \quad 0.632K_0 \quad 1\|'$$

$$^+)=\|B\|\mathbf{v}(T)$$

$$=\|1 \quad 0.358K_0 \quad 0.632K_0 \quad (1-0.368K_0)\|'$$

ilarly, for $n=1$

$$T)=\|\phi_1\|\mathbf{v}(T^+)$$

$$= \begin{Vmatrix} 1 \\ 0.768K_0+0.368(1-0.368K_0)K_1 \\ 0.232K_0+0.632(1-0.368K_0)K_1 \\ 1-0.368K_1 \end{Vmatrix}$$

or the system to perform with dead- response, the following conditions t be fulfilled:

$$T)=0.768K_0+0.368(1-0.368K_0)K_1=1$$

$$T)=0.232K_0+0.632(1-0.368K_0)K_1=0$$

olving these two simultaneous equa- s gives $K_0=1.58$ and $K_1=-1.38$; , $m(T^+)=1-0.368K_0=0.418$.

he input and output of the D-N bination are

$$)=1+0.418z^{-1}$$

$$)=1.58-0.577z^{-1}$$

aking use of Fig. 3, the input to the inear element is found to be $M_1(z)=2-0.636z^{-1}$. Hence the pulse-trans- unction of the desired digital con- er is given by

$$=\frac{1.092(1-0.582z^{-1})}{(1+0.418z^{-1})}$$

ne output of the compensated system otted in Fig. 4, which reaches steady- without overshoot in two sampling ods.

MPLE 2, SATURATION ONLINEARITY

nsider the nonlinear sampled-data rol system of Fig. 1(A) with $G(s)=1/s(s+1)$ and nonlinearity shown in Fig. 5(A). The sampling period is assumed to be one second, and the input is a step function of two units. This is the same example considered in reference 9.

The transition matrix and the B-matrix have been found in example 1. For $n=0$

$$\mathbf{v}(0^+)=\|B\|\mathbf{v}(0)=\|2 \quad 0 \quad 0 \quad 2\|'$$

and

$$m(0^+)=2$$

Since

$$m(0)^+>1,\ m_2(0^+)=1 \text{ and } K_0=0.5$$

Thus

$$\|\phi_0\| = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0.632 & 0.184 \\ 0 & 0 & 0.368 & 0.316 \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

$$\mathbf{v}(T)=\|2 \quad 0.368 \quad 0.632 \quad 2\|'$$

$$\mathbf{v}(T^+)=\|2 \quad 0.368 \quad 0.632 \quad 1.632\|'$$

which gives $m(T^+)=1.632$. Since $m(T^+)>1$, $m_2(T^+)=1$ and $K_1=0.612$. Hence

$$\|\phi_1\| = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0.632 & 0.225 \\ 0 & 0 & 0.368 & 0.387 \\ 0 & 0 & 0 & 1 \end{Vmatrix}$$

For $n=1$

$$\mathbf{v}(2T)=\|2 \quad 1.135 \quad 0.864 \quad 2\|'$$

$$\mathbf{v}(2T^+)=\|2 \quad 1.135 \quad 0.864 \quad 0.865\|'$$

$$m(2T^+)=0.865$$

Since $m(2T^+)$ is less than one, K_2 must be determined from equations 21 and 22. For $n=2$

$$\mathbf{v}(3T)=\begin{Vmatrix} 2 \\ 1.681+0.318K_2 \\ 0.318+0.547K_2 \\ 0.87 \end{Vmatrix}$$

and

$$\mathbf{v}(3T^+)=\begin{Vmatrix} 2 \\ 1.681+0.318K_2 \\ 0.318+0.547K_2 \\ 0.319-0.318K_2 \end{Vmatrix}$$

Similarly, for $n=3$

$$\mathbf{v}(4T) = \begin{Vmatrix} 2 \\ 1.882+0.664K_2+0.117(1-K_2)K_3 \\ 0.117+0.201K_2+0.201(1-K_2)K_3 \\ 0.319-0.318K_2 \end{Vmatrix}$$

Conditions for deadbeat performance are

$$1.882+0.664K_2+0.117(1-K_2)K_3=2$$

$$0.117+0.201K_2+0.201(1-K_2)K_3=0$$

These two simultaneous equations have the solution

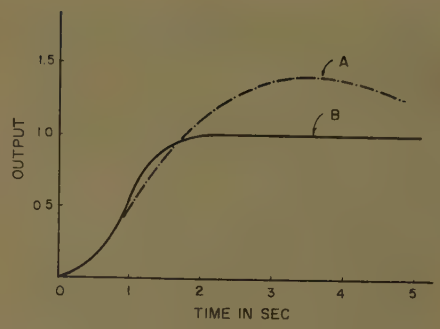


Fig. 4. Step-function response of example 1

A—Uncompensated
B—Compensated

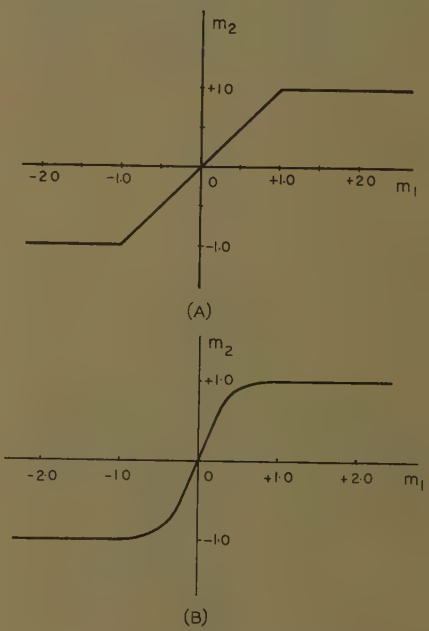


Fig. 5. A—Idealized saturation curve. B—Saturation curve

$$K_2=0.341 \text{ and } K_3=-1.41$$

Hence, the input and output of the D-N combination are

$$M(z)=2+1.632z^{-1}+0.866z^{-2}+0.210z^{-3}$$

$$M_2(z)=1+z^{-1}+0.295z^{-2}-0.296z^{-3}$$

Since the element N is linear between the saturation limits, $M_1(z)=M_2(z)$. Therefore

$$D(z)=\frac{0.5(1+z^{-1}+0.295z^{-1}-0.296z^{-3})}{(1+0.816z^{-1}+0.433z^{-1}+0.105z^{-3})}$$

The step-function response of the compensated system is plotted in Fig. 6, which reaches the steady-state without overshoot in four sampling periods. This checks with the results obtained by Mullin, but the foregoing synthesis procedure is simpler and more systematic.

Now, consider a nonlinear element characterized by the more realistic curve of Fig. 5(B), which has the same saturation

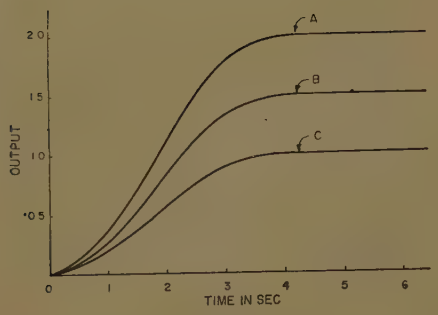


Fig. 6. Step-function response of example 2. Unit step inputs: A—2.0. B—1.5. C—1.0

limits as the idealized curve of Fig. 5(A) but is nonlinear between the saturation limits. The input and output of the *D-N* combination remain unchanged. But the input to the nonlinear element is changed to

$$M_1(z)=1+z^{-1}+0.1z^{-3}$$

Therefore the required pulse-transfer function is

$$D(z)=\frac{0.5(1+z^{-1}+0.1z^{-2}-0.1z^{-3})}{(1+0.816z^{-1}+0.433z^{-2}+0.105z^{-3})}$$

The output of the compensated system in response to the specified step-function input is the same as the system with idealized saturation nonlinearity. It is observed that the deadbeat response depends upon the magnitude of the step input and the saturation limit; however, the nonlinear characteristic between the saturation limits has no influence upon the output of the compensated system.

Example 3, Linear System

The synthesis procedure developed in this paper can be applied to linear systems as well. As an illustration, the system of example 10.3-1 in reference 1 is designed by this variable-gain approach. The plant of this system has a transfer function $G(s)=10/s(s+1)$. The system, incorporated with a zero-order hold, samples with sampling period equal to one second.

With the desired digital controller $D(z)$ represented by a variable gain K_N, one derives the transition matrix and the B-matrix by inspection of the state diagram of this system:

$$\|\phi(T)\|=\begin{Vmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0.632 & 3.68K_n\\ 0 & 0 & 0.368 & 6.32K_n\\ 0 & 0 & 0 & 1\end{Vmatrix}$$

$$\|B\|=\begin{Vmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 1 & -1 & 0 & 0\end{Vmatrix}$$

For $n=0$

$$\mathbf{v}(T)=\|1 \quad 3.68K_0 \quad 6.32K_0 \quad 1\|'$$

$$\mathbf{v}(T^+)=\|1 \quad 3.68K_0 \quad 6.32K_0 \quad 1-3.68K_0\|'$$

For $n=1$,

$$\mathbf{v}(2T)=\begin{Vmatrix}1\\ 7.68K_0+3.68(1-3.68K_0)K_1\\ 2.32K_0+6.32(1-3.68K_0)K_1\\ 1-3.68K_0\end{Vmatrix}$$

For the system to meet the specified requirements, the following conditions must be fulfilled:

$$7.68K_0+3.68(1-3.68K_0)K_1=1$$

$$2.32K_0+6.32(1-3.68K_0)K_1=0$$

The solutions of these two simultaneous equations are $K_0=0.158$ and $K_1=-0.138$.

Thus

$$m(T^+)=1-3.68K_0=0.418$$

$$m_1(0^+)=0.158 \qquad m_1(T^+)=-0.0578$$

The desired pulse-transfer function is

$$D(z)=\frac{0.158(1-0.368z^{-1})}{(1+0.418z^{-1})}$$

This checks with the result of reference 1.

The above method of synthesis provides an alternate approach; it, again, appears simpler and more systematic.

Conclusions

The development of the design technique described in this paper is based upon the variable-gain concept. The digital controller and the nonlinear element are represented by a variable gain K_N in the state diagram of the system. In carrying out the design, use is made of the powerful state-transition technique. This synthesis procedure is simple and systematic, involving four major steps which can be programmed on a digital computer for the design of optimum control.

As illustrated in the preceding section, this technique also provides a simpler and more systematic approach to the synthesis of linear discrete-data control systems. Furthermore, this unified approach can also be used to synthesize nonlinear multirate, variable-rate, and finite-pulsed systems, and to design nonlinear control for the compensation of linear and nonlinear systems. Although only the deadbeat-response criterion has been considered, the variable gain approach and the synthesis procedure may be extended to the design of nonlinear control systems with other specified performance criteria.

References

1. Digital and Sampled-Data Control Systems (book), J. T. Tou. McGraw-Hill Book Company, Inc., New York, N. Y., 1959.

2. Sampled-Data Control Systems (book), J. R. Ragazzini, G. F. Franklin. McGraw-Hill Book Company, Inc., 1958.

3. Sampled-Data Control Systems (book), E. I. Jury. John Wiley & Sons, Inc., New York, N. Y., 1958.

4. Analysis of Sampled-Data Control Systems Containing a Nonlinear Element, J. T. Tou. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 46, May 1958, p. 915.

5. Analysis of Nonlinear Sampled-Data Control Systems—I, E. Kinnen, J. Tou. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 79, Jan. 1960, pp. 386–90.

6. Analysis of Nonlinear Sampled-Data Control Systems—II, *Ibid.*, pp. 390–94.

7. Contactor Servomechanisms Employing Sampled Data, C. K. Chow. *Ibid.*, vol. 73, 1954, pp. 51–64.

8. Nonlinear Aspects of Sampled-Data Control Systems, R. E. Kalman. *Proceedings*, Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Interscience Publishers, Inc., New York, N. Y., vol. 6, 1957, pp. 273–83.

9. The Stability and Compensation of Sampled-Data Systems, Francis J. Murray. *AIEE Transactions*, pt. I (*Communication and Electronics*), vol. 78, July 1959, p. 270.

10. Solution of Variable Circuits by Matrices, L. A. Pipes. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 224, Dec. 1937, pp. 77.

11. A Unified Approach to the Theory of Sampling Systems, R. E. Kalman, E. Bertram. *Ibid.*, vol. 267, 1959, p. 405.

12. A Method for the Symbolic Representation and Analysis of Linear Periodic Feedback Systems, Edward O. Gilbert. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 79, 1960, pp. 512–23.

Discussion

H. C. Torng and W. E. Meserve (Cornell University, Ithaca, N. Y.): The authors are to be commended for their lucid presentation of a systematic and powerful technique for the synthesis of nonlinear discrete-data control systems by using the concept of variable gain.

We would like to point out that, in synthesizing a linear discrete-data system to have a deadbeat performance, the designer can determine the settling time by simply examining the number of zeroes of the pulse transfer function of the plant and the characteristic of the test input. It seems to us that this will not be the case in a nonlinear situation.

The simultaneous equations to be solved to determine the variable gains are linear. We would like to know whether solutions with real values are guaranteed in each case, because any solution involving complex numbers will render this approach useless; or, is there a solution at all?

In using this technique, the designer needs to be alert to check a possible solution at each step because any delay will mean a more costly compensator.

As the examples in the paper are for a cond-order plant, only two equations are volved. In a higher order plant, the task finding the solution of a set of nonlinear nultaneous equations is quite formidable. e would like to know whether the authors ve tried any higher order systems.
In general, however, this is a promising proach which should undergo further in-stigation.

ilius T. Tou: We wish to thank Professors rng and Meserve for their discussion. ey are correct when they say that the signer cannot determine the settling time simply examining the transfer function the linear portion of the plant. In fact, e settling time is dependent upon the ture of the nonlinearity preceding the ear plant. Saturation nonlinearity will ually cause a longer settling time because the limited amount of energy contained the control sequence. As illustrated by the three examples in the paper, for the second-order linear system the step-function response settles in two sampling periods, as expected. When the nonlinearity is of the nature shown in Fig. 2, the response also settles in two sampling periods. However, when the saturation nonlinearity is present, the response will settle in four sampling periods. Although there is no simple way to determine the settling time by inspection of the plant transfer function, the settling time can readily be found from the given nonlinearity and the transition matrix in the process of determining the variable gain. The settling time is given by the number of variable gain constants K_n to be determined.

It is quite correct that the simultaneous equations to be solved to determine the variable gains are not linear. However, it should be noted that these equations are always in forms that can easily be reduced to linear expressions through elimination, as illustrated by the three examples in the paper. It is true that for a physical system the variable gains must be real numbers. Any solution involving complex or imaginary numbers does not make this approach useless, but it means that the deadbeat performance is unattainable with the given control scheme. Since this is a synthesis technique and the solution is unique, we do not think the designer has to worry too much about any delay in carrying out the synthesis procedure. As to the last question, we feel that the best way to clarify theoretical developments is to illustrate by simple examples. That is why only second-order plant has been considered in this paper. If fact, we have tried successfully to design the digital controller for a nonlinear system with third-order plant.

In conclusion we wish to emphasize the fact that although the paper is entitled "Optimum Control of Nonlinear Discrete-Data Systems," the technique proposed actually provides a novel way of designing nonlinear feedback control systems.

Waveshape Effect on Alloying and Arc Stability of A-C Tungsten Inert-Arc Welding

THOMAS B. CORREY
MEMBER AIEE

ALUMINUM-CLAD aluminum-silicon-alloy bonded nuclear fuel elements are fusion-welded over the exposed ized joint to assure a watertight clo-e. As is common with some brazed nts, there are porosities and pipes in braze which may provide a path for cooling water to attack the uranium tal. Work with the fusion welds disclosed that the braze metal was alloying with the cladding into a nogeneous weld alloy. In many in-nces the braze metal was continuous m the parent braze metal to the surface the weld. Within limits of normal ding speeds but a very small change in ying occurred with changing welding ed. Current programming was de-ble to improve weld quality. This trolled the weld width as the weld gressed and raised the temperature of metal to be melted. The welding ver supply in use had a manually oper-d mechanical current control which ld not be readily converted to current gramming.

A welding power supply with a remote tric current control was tried. The technique used for the original power supply was applied, and the alloying which resulted was very inferior. Alloying equal to the original unit could not be produced with any technique.

From current waveshape and arc voltage studies of the two units, it was found that the original unit produced a sine wave of current, and the new unit a sine wave with a 180-degree out-of-phase third harmonic. In addition, the current was not stable with the latter unit. A study of the power half-cycles of each unit indicated a much longer period between the power pulses for the out-of-phase third harmonic sine current wave than for the sine wave. The difference in the spacing of the power pulses indicated that the temperature gradient in the welds was much higher with the out-of-phase third harmonic current wave. From this it was concluded that if the temperature gradient in the weld between half-cycles could be held at a more constant value, the weld alloying could be improved.

In a study of current waveshapes from the out-of-phase third harmonic to a square wave, it was found that as the waveshape changes toward a square wave, the alloying improves. Also, for the same rms sine wave of open-circuit voltage and argon shielding gas, the out-of-phase third harmonic current wave required continuous superimposed high frequency to maintain the arc with the alternate half-cycles of current not of a uniform value, while the square wave maintained a uniform current without superimposed high frequency.

Electric Arc Phenomena

The electric arc is a complex phenomenon. In arc welding very few of the phenomena do not affect the welding and very few are not of major interest.[1] Arc welding of all types employs a high-pressure arc, starting essentially at a point

Paper 61-546, recommended by the AIEE Electric Welding Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE-AWS Electric Welding Conference, New York, N. Y., April 17–21, 1961. Manuscript submitted August 8, 1960; made available for printing February 17, 1961.

THOMAS B. CORREY is with the General Electric Company, Richland, Wash.

The author expresses his thanks to Oregon State College for permission to use this material, which was originally submitted as a thesis for the degree of Electrical Engineer, June 6, 1960. Many persons have contributed to the success of this paper; the author wishes to acknowledge the contribution of Louis N. Stone, Head, Electrical Engineering Department, Oregon State College, for encouragement and counseling; Kirk Drumheller, of the General Electric Company, Fuels Preparation Department, Richland, Wash., for his recognition of the possibilities in the study, and his procuring the required funds; and Engineering Assistant, D. E. DeWitt, for making and preparing the many thousands of welds and weld sections necessary in evaluating the various current wave shapes; to all others, regardless of the extent of their contribution, the author expresses his sincerest thanks.

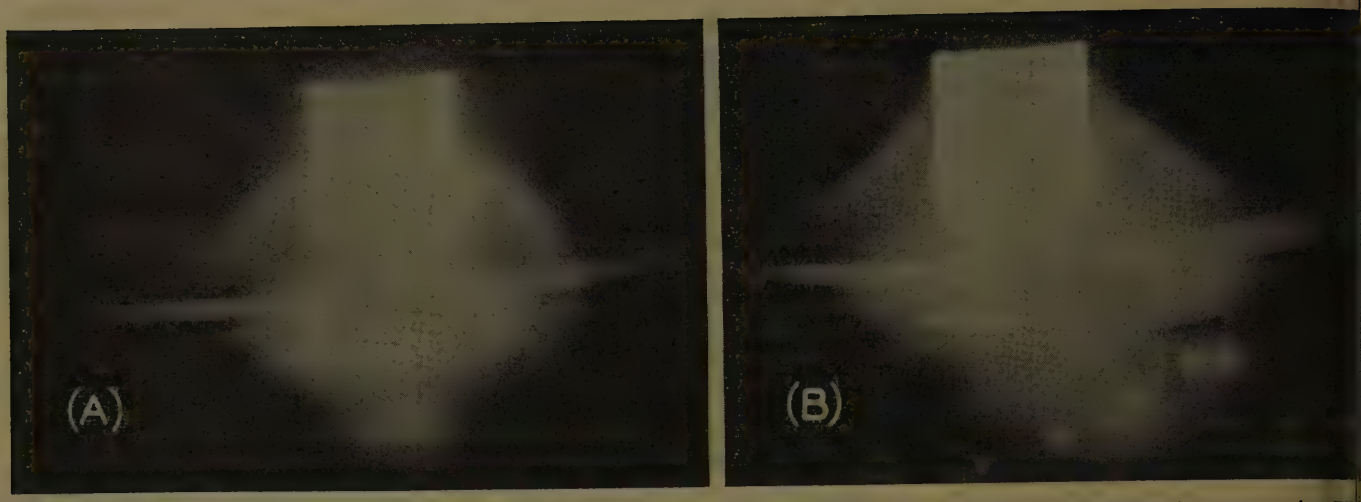


Fig. 1. Arc shapes 16-5amp, 5/32-inch zirconium-tungsten electrode, 5-to-7 argon-helium volume ratio and balanced current. A—Electrode negative. B—Electrode positive

and ending on a plane. The point-to-plane arc is a plasma jet, the velocity of which is determined by the diameter of the electrode, irrespective of the polarity or electrode material.[2] The polarity of the point-to-plane arc determines its shape and thus the heat concentration on the plane or work piece.

Velocity of the plasma jet is inversely proportional to the electrode diameter; it determines the penetration of the weld for all types of arcs and produces the metal transfer and its characteristics in the consumable electrode arc. The velocity of the jet is on the order of 10^4 cm/sec (centimeters per second).

The temperature of the high-pressure tungsten arc varies from 30,000 K (degrees Kelvin) at the core to 5,000 K in the outer regions.

With the electrode negative and the work piece positive, the arc has a hemispherical shape, Fig. 1 (A), producing a small, intensely heated anode spot on the work piece from electron condensation.[3] Virtually all of the electrons are supplied from the cathode spot on the electrode, a very few being supplied by the positive ion streamers, metal vapors, and shielding gas adjacent to the cathode. Etching of the electrode shank indicates a sheath of positive ion streamers around the main arc core.

The negative end of the arc first establishes a cathode spot on the work piece by positive ion bombardment. This form of the arc is composed of a primary cathode spot and many secondary spots that are constantly forming and vanishing (see reference 1, p. 87, and reference 4). These are shown as the bright spots in Fig. 1 (B), taken from a high-speed motion picture of an a-c arc. A very intense primary cathode spot travels over the surface of the work piece with a velocity on the order of a 10^4 cm/sec (see reference 1, p. 87). During this phase of the arc the current is carried primarily by positive ions. This phenomenon with the tungsten inert arc is the basis for a new patent for cleaning metal surfaces.[5]

As soon as enough heat is added to a finite spot on the work piece to produce melting, the primary arc anchors to this spot, the voltage drops, and the current is virtually an electron current. The electrons necessary to maintain the arc are supplied mainly by the metal vapors, but partially by the shielding gas adjacent to the cathode and the action of the positive ion streamers on the oxide interface. The arc has a cone shape producing a broad, low-intensity heated spot; see Fig. 1 (B). It is known that the temperature of the molten metal in welding is not high enough to supply by thermal electron emission the electrons required by the true arc.[6] The positive ion portion of the current has been estimated at less than 1.0% of the total current (see reference 1, p. 75). When the arc is in this form, the many secondary spots continue to form and vanish.

The shape and thus the heat concentration of the tungsten arc with constant current is controlled by the shielding gas used. As the mixture of gas changes from pure argon toward pure helium, the arc becomes more concentrated and the arc voltage increases.[7] The end effect is increased penetration for equal currents. Lengthening of the arc increases its total power but does not increase the penetration (see reference 1, p. 85).

With cathodic etching as a cleaning method, recent work indicates that it is difficult for the secondary cathode spots to pierce heavy oxide layers but, once through, they have an affinity for the interface between the oxide layer a the metal. Etched areas appear 1 in each side of the weld center line wh Zircaloy-2 and copper are welded wi 200-amp (ampere) balanced square-wa alternating current. In high-speed m tion pictures the streamers have bl observed returning to the same a three times and removing a lay The width of the etched area each s of the weld is a function of the metal increases from aluminum to copper Zircaloy-2. There is no evidence to dicate that a nonetched area is ever n closed in an etched area during weldi The oxides act as a thermal barrier reference 4, pp. 15–22).

In welding such metals as zirconi copper, and their alloys with alterna current, the positive ion bombardm during the reverse polarity half-cy is very great, but good welds can be duced. Third harmonic distorted current waves are better for welding per and zircaloy, as the effective pe during the half-cycle in which the ar ignited is approximately one half the t cycle period. Also the etching de as the current starts to decay. Tw two factors limit the distance from center line of the weld over which cathode spot is able to travel. If t time period of the reverse polarity could be made such a small portion of total cycle time that it just cleaned weld area, then it is believed welds superior to d-c straight-pola welds could be produced. This me of welding and the equipment for pro ing it are the subject of a United St Atomic Energy Commission patent plication.

The electrical characteristics of a stant-length arc in a particular gas with a finite current are determine

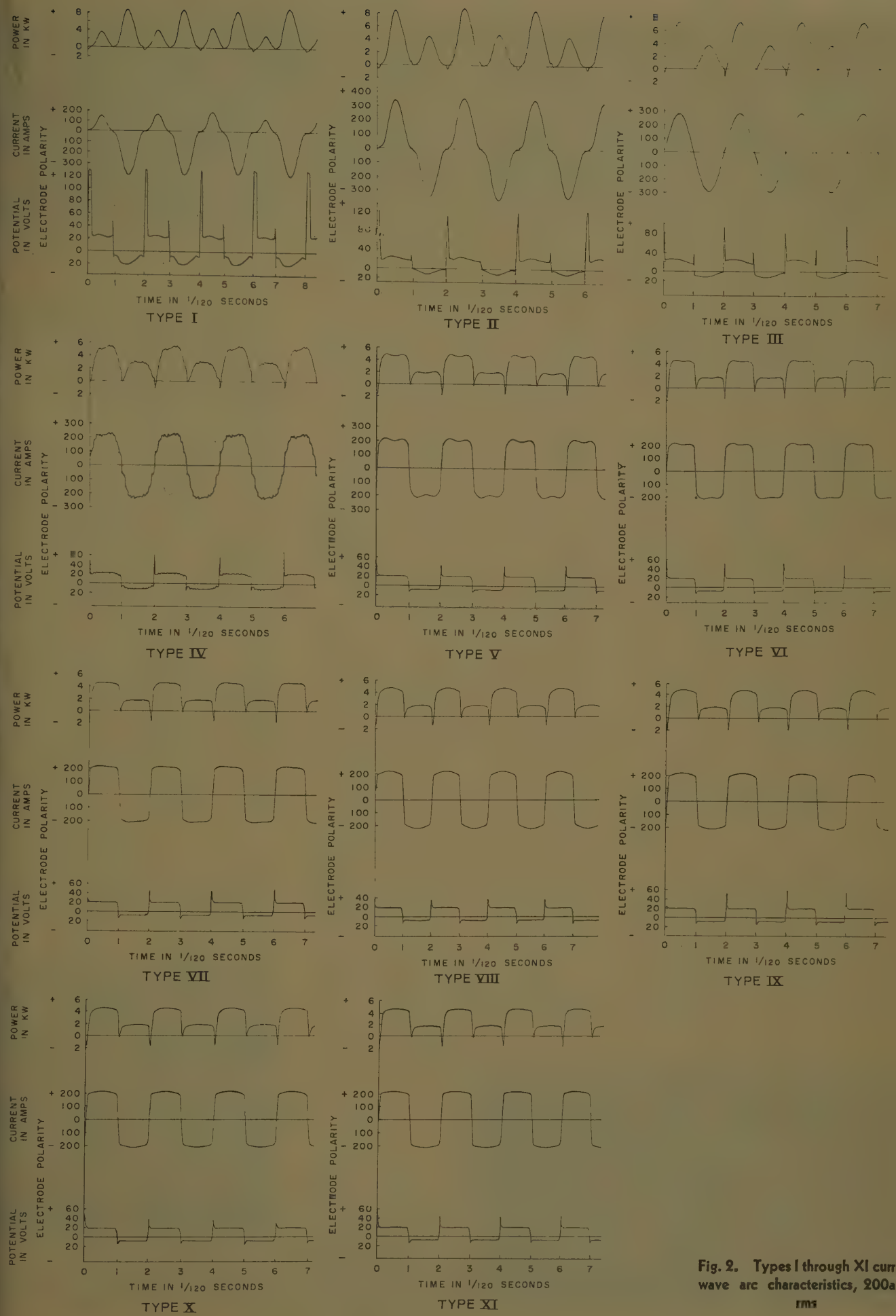


Fig. 2. Types I through XI current wave arc characteristics, 200amp rms

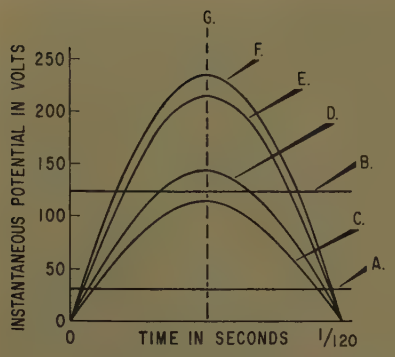


Fig. 3. Transformer voltage characteristics, 80 open-circuit volts, 1/8-inch tungsten electrode, 165 amp; argon shielding gas. Maximum voltage required to restrike arc with tungsten electrode (A) negative; (B) positive. Volts rms: (C) 80; (D) 100; (E) 150; (F) 165; (G) Approximate arc restrike point

the electrode materials, the relative electrode dimensions, and the polarity in relation to electrode dimensions. If the electrode is tungsten and the work piece is aluminum with a shielding gas of five parts by volume of argon to seven of helium for a capacitor-balanced square wave of current at 200-amp rms, the arc drop will be 8 volts with the electrode negative and the work piece positive; with the electrode positive and the work piece negative the arc drop will be 20 volts. If the capacitors are removed, the current during the half-cycle in which the electrode is positive will be small compared to that when the electrode is negative (see Fig. 2, type I). This phenomenon is called rectification in welding; it occurs because the tungsten electrode is a thermal emitter and such metals as aluminum are a plasma or field emitters (see reference 1, p. 87).

Rectification is ever present in arc welding with alternating current, even though the electrode and work piece are of the same material; the difference in physical dimensions will affect the electron emission relationship between them.[8] The use of series capacitors in the welding current circuit is the commonest of the three methods of eliminating rectification.[9] Rectification has two effects on welding and its associated equipment: the first is the reduction of energy in the half-cycle when the electrode is positive; the second is the saturation effect on the magnetic cores in the welding power supply. This is particularly evident in the saturable reactor type of unit.[10]

In an electric arc the heat generated at the positive end is approximately twice that generated at the negative end when the negative electrode is a refractory material (see reference 8, p. 521). In addition, the unit energy density at the cathode, excluding the cathode spot, is very low because of the large area covered. Therefore, in the case of the tungsten a-c arc in which there is rectification, the half-cycle during which the work piece is negative is essentially eliminated as a heat source.

For a very brief period at the end of each half-cycle in an a-c arc, the current may be zero, depending on the rate of rise of the reversing voltage and the external circuit constants. During this period the ionization decreases, which decreases the conductivity of the arc column. In addition, the temperature of the new cathode drops rapidly. As the voltage reverses, the arc conductivity that existed just before the instant of zero current must be re-established. Also, the polarities of the space charges that exist at the old anode and cathode must be reversed. The voltage to reignite the arc is consequently higher than the arc-burning voltage.

The process of reignition may be considered as a race between the rising recovery voltage and the forces of deionization and temperature drop.[11,12] The potential drop at the start of the reignition period is largely concentrated in the space adjacent to the new cathode, the major portion of the arc being free of potential gradient.[13] In a circuit with a high ratio of inductance to resistance, arc reignition is improved, for at zero current the voltage approaches a maximum, and the distributed capacitance can cause the recovery voltage to reach a peak value double that of the open-circuit voltage (see reference 14 and p. 926 of reference 13). This rise of voltage is shown in Fig. 2, type I, in which the restrike potential is 125 volts for an 80-volt rms sine wave open-circuit potential which produces a maximum potential of 113 volts. The use of high values of inductance that produce time constants of more than 1/2 sec adversely affect the programming of short weld cycles.

As the recovery voltage reverses and the current passes through zero, the forces of deionization, temperature drop, neutralization, and reversal of the space charges at the electrode surfaces cause the arc to re-establish always as a glow discharge requiring a higher arc voltage than for the true arc. Transition from a glow discharge to true arc occurs when sufficient electron emission is established. At the transition from the glow discharge to the true arc, the arc voltage drops suddenly to that required for the true arc (see reference 3, p. 355).

In welding with the tungsten inert-gas shielded arc, when the work piece is aluminum with its high thermal conductivity, the problem of arc reignition becomes very critical. The arc restrikes at a high voltage when the aluminum is positive because of the high thermal conductivity and low thermal electron emission characteristic of aluminum.

The high value of restrike voltage with its wide separation when the electrode becomes positive is shown in Fig. 2, type I. Fig. 3 shows a family of curves taken on an unbalanced current unit which has an open-circuit voltage, i.e., the voltage before the arc is ignited, which is sinusoidal, as well as a sinusoidal welding current, type III. A pure tungsten electrode, an aluminum work piece, and argon shielding gas were used. In this case the 100 volts were not enough to assure positive arc reignition and 150 volts were just on the safe side. In addition, as the welding current approaches zero, a similar condition exists and a glow discharge will form if the recovery voltage and rate of rise are not high enough or if the cooling and deionization rates are too high. If the voltage is not high enough to equal or exceed the voltage to restrike the arc when the electrode is positive, the arc will either go out or conduct only when the electrode is negative (see reference p. 87).

Increasing current normally produces decreasing arc-burning and restrike vol-

Fig. 4. Weld section with desirable alloying, enlarged 20×, caustic etch

Fig. 5. Closures, enlarged 1¼×. A—Welded. B—Unwelded

ges within the current limits of a particu-
ar electrode. Near and above the upper
urrent limit there is a slight rise in the
otal arc voltage. As the electrode
iameter is increased for a constant value
f current, the arc-burning and restrike
oltages increase. This also applies to
ncreasing arc length. With a 100-volt
pen-circuit) capacitor balanced sine
ave of current, 1/8-inch tungsten
lectrode, aluminum weldment and argon
hielding gas, increasing the current from
0 to 170 amp decreases the arc restrike
oltage one half and the burning voltage
ne third. The arc restrike and ex-
inguishing voltages decrease as the
urrent wave squares up, the open-
ircuit voltage increases, and the welding
peed decreases. For the 180-degree
ut-of-phase third-harmonic distorted
apacitor-balanced current wave, 165
pen-circuit volts are required for the
ame welding conditions as for a capacitor-
alanced square wave and 100 open-
ircuit volts.

usion Weld Closures

From a corrosion and dependability
tandpoint, it is desirable to produce a
weld completely alloyed in all its parts, thus assuring that it will be free of voids, pipes, and cracks. Shown in Fig. 4 is a weld section with the desired alloying. Also, fine surface smoothness and uniform weld width are required. A pair of welds removed from the ends of a randomly selected single fuel element is shown in Fig. 5(A).

To obtain a uniform base on which to determine weld quality, the width across the end, the overhang over the cladding wall, and the depth down the cladding wall were made the same on all welds.

In Fig. 5(B) is shown the unwelded type of male and female closures of the AlSi-bonded aluminum-clad uranium-metal fuel element used in the study. A section through the zone to be fusion-welded is shown in Fig. 6.

ig. 6. Unwelded closure sections: at left, emale end; at right, male end. Enlarged 10×; caustic etch

Welding Gases

A 5/7 argon-helium volume ratio was used for all the tests, with the same flow rate maintained for all welds. This ratio was selected on the basis of weld-alloying improvement, depth to width ratio increase, cleanliness of the weld surface, and increased welding speed as compared to argon.

Welding Current

The welding machines and power supplies available for performing the tests were developed over a period of years and varied from a semiautomatic machine to fully automatic machines. In the former, the operator placed the work piece in a fixture and pressed the button to initiate the welding cycle; in the latter the work pieces are brought in for welding and taken out after welding by conveyor. The semiautomatics were originally designed without current programming and had a current decay at the end of the weld. These units were subsequently modified to produce the current programs shown in Fig. 7(A). To produce smooth arc starting, a weld of uniform dimensions, and a closure point of the same dimension as the weld, the fully automatic machines were designed with seven steps of current programming; see Fig. 7(B).

The current programs for current wave types I and II, Fig. 8, started at full welding current; near the closure point of the weld an air-operated device was actuated to decay the current rapidly. The type III current wave programs are shown in Fig. 7(A), and those for types IV through XI, in Fig. 7(B).

Current Waveshapes

Previous work has established that there is definitely a relationship between the alloying produced in a fusion weld closure of an AlSi-bonded aluminum-clad uranium-metal fuel element and the current waveshape of the alternating current used to produce it (see references 4, pp. 53–55, and 10, p. 3). To obtain as broad a sampling as possible, current waveshapes from all but two types of welding power supplies tested were used

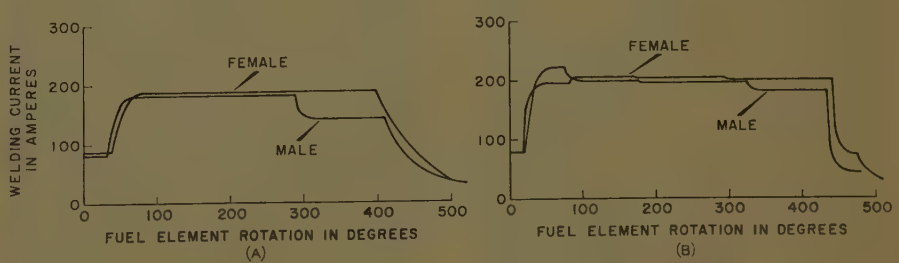


Fig. 7. Welding machine current programs. A—Semiautomatic. B—Fully automatic

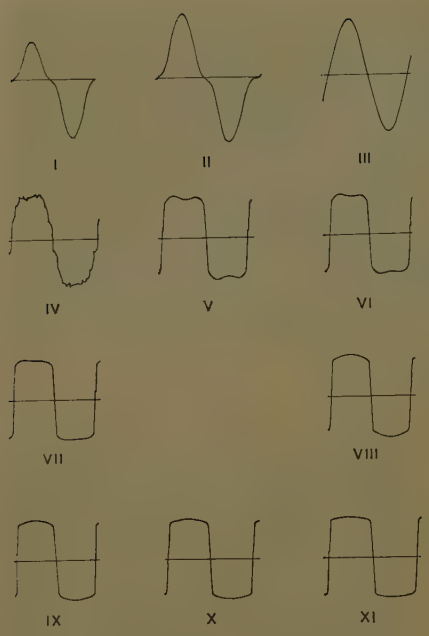


Fig. 8. Current waveshapes investigated

for comparison. Types I and II were not used in this test as previous work indicated that they were completely unsatisfactory (see reference 4, pp. 41–46). The types of current waveshapes used in this and the previous tests are shown in Fig. 8.

The current waves in Fig. 8 were produced for:

Type I, by the commercially available saturable-reactor type of units without current wave balancing.

Type II, by using series current balancing capacitors in the welding current circuit of the unit used for type I.

Type III, by the commercially available movable coil type units with series current balancing capacitors.

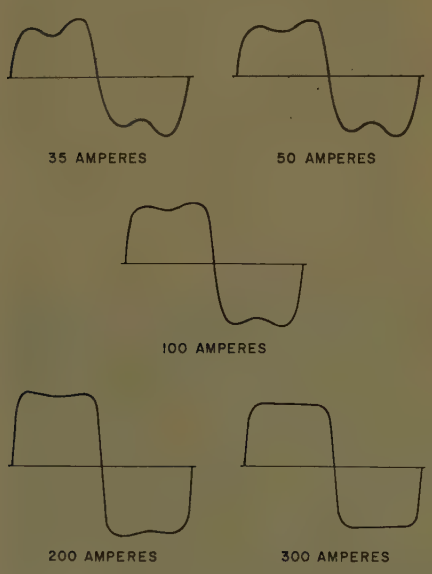


Fig. 9. Type VII current waveshape variation with current

Types IV through XI, by custom designed and fabricated units with series current wave balancing capacitors.

For all the saturable-reactor type of devices the current waveshape appears to change as the current changes from minimum to maximum or vice versa. The waveshape that occurs at minimum current becomes the crest of the wave as the current is increased. In Fig. 9 is shown the variation in current waveshape of a wave similar to the type VII wave with the current varying from minimum to maximum. If all of the ordinates were at the same scale, the exact ordinates of the 35-amp wave would appear on the crest of the 300-amp wave. This means that as the amplitude of a square wave increases, the form factor approaches unity. The maximum current waveshape variation can be reduced by using either shorter current ranges or multiple current ranges. The current range below 100 amp was used in arc starting and crater filling and it produced no detectable effect on weld quality.

Weld Classification

The following weld classification system was developed for the purpose of placing a numerical value on the weld quality of a sample lot of 100 welded closures from one end of a fuel element. The closures are sectioned through the end of the crater-filling period, polished with *240X* grit belt, and caustic etched. This produces 400 welds for classification purposes.

To each class is assigned a digit equal to the class number. The largest digit, 5, was assigned to class V welds so that a slight change in the number of the poorest welds would produce the greatest change in the weld quality number. When the reading of the total weld section surfaces is completed, the total of each class is multiplied by the assigned digit. These totals are added and the new total is divided by the number of classes. The division by the number of classes reduces the weld quality number to a smaller, more easily handled number. The quotient produced is a measure of weld quality; the smallest number possible being 80 for all perfect welds, and the largest number 400 for all bad welds. In all the welds there must be no evidence of reduction of the residual cladding thickness by the welding. Descriptions of the various classes shown in Fig. 10 are listed.

Class I—Excellent. A weld in which the alloying throughout the weld section is complete and there is no evidence of localized silicon concentration in the weld or heat affected zone.

Class II—Good. A weld in which there is complete alloying in the weld from the braze to the surface of the weld equal to the thickness of the residual can or tube cladding. There may be points of localized silicon concentration as long as the incremental in line total of complete alloying is equal to the residual cladding thickness.

Class III—Fair. The same as class II except that the incremental in line total must be at least one half the residual cladding thickness.

Class IV—Poor. The same as class III except that the incremental in line total must be less than one half of the residual cladding and must be greater than 0.000 inch. In addition, this class includes all welds in which there is a continuous path of decreasing silicon concentration from the braze to the surface of the weld. This path must become invisible at the surface of the weld.

Class V—Bad. A weld in which there are (1) massive areas of silicon concentration which are continuous from the braze to the surface; (2) continuous path of silicon concentration from the braze to the weld surface; or (3) in which items 1 and 2 come to 0.005 inch or less of the weld surface.

Class VI—Internal cracks. Cracks occurring in the interior of the weld mainly at the closure point, and not visible on the weld surface.

Class VII—Spots. Localized high-silicon spots occurring on the surface of the weld and extending to the braze in a continuous column.

Class VIII—High-silicon quarters. The spots are so close together that it is impractical to count them. They are classified as nearest fourth of the circumferential weld distance they cover.

Classes VI and VII are not included in the calculations as their frequency of occurrence is very erratic compared to other classes. Class VIII was not included as it occurred only in the second group of material.

Current Waveshape Evaluation

The very complex assembly process of AlSi-bonded aluminum-clad uranium metal fuel elements is classified as confidential Atomic Energy Commission information. Many factors affect the quality and analysis of the braze. In canning a group of materials for finite test it is virtually impossible to control all of the factors affecting the particular condition studied. Statisticians, after intensive study of the assembly process conclude that no real meaning results from applying statistics to the findings of a welding test.

For reasons beyond control, it was not

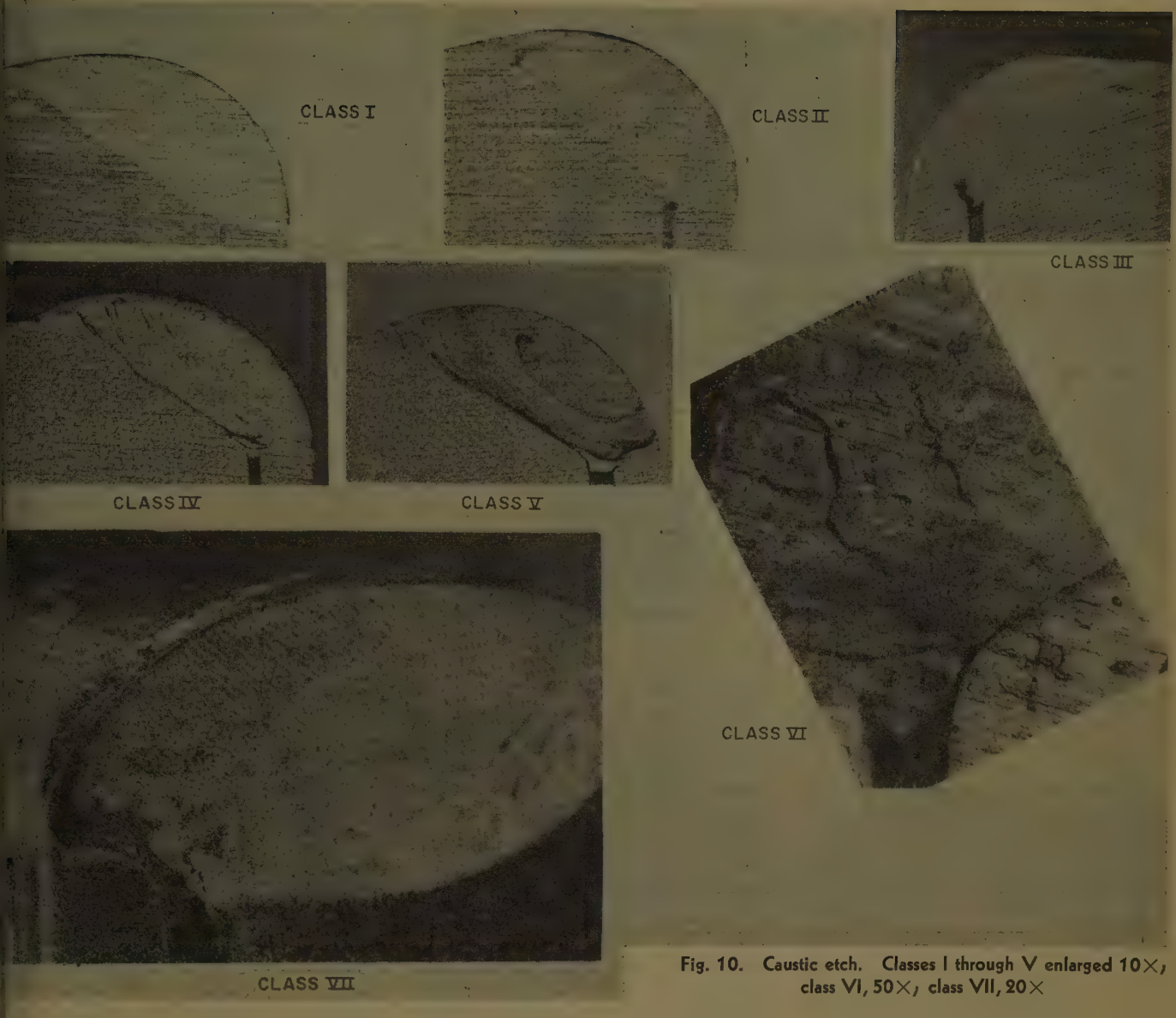


Fig. 10. Caustic etch. Classes I through V enlarged 10×, class VI, 50×, class VII, 20×

ossible to obtain enough identical uran-
m cores for testing so that all of the
quired pieces could be canned at the
me time with the same equipment and
ews. In addition to the finished
aterial required, enough uranium cores
ust be started through the preparation
ocess to take care of the shrinking
curring at preparation, canning, and
cing. For these reasons the decision
as made to can material for the tests in
o groups.

To complicate the problem further,
ough how extensively was not known
the time, the cladding material had
en changed from *1245* alloy, which is
sentially pure aluminum, to *X-8001*,
hich contains approximately 1.0% nickel.
was known at this time that the addi-
n of nickel to the aluminum made the
oying in the weld more difficult but
thing was known of the effect of
rying concentrations. With the *1245*
dding all of the machining scraps were converted into *A-13* brazing metal. This practice was continued with the advent of the *X-8001* cladding material, which increased the nickel content in the braze metal. The first canning group of material produced excellent welds and normal weld quality correlation with known current waveshapes.

Table I shows the variation in quality of the mating surfaces of a group of type IV current wave weld sections from canning no. 2, indicating that the weld quality correlation around a single weld and throughout a number of welds is not good and that the male weld is more difficult to produce than the female. Throughout the second canning group there was no complete over-all weld quality correlation, and some of the data reversed known relationships.

Reading of the weld sections disclosed a difference in the aluminum grain size and in the darkness of the silicon matrix. The matrices were uniform throughout each of the two cannings. Fig. 11(A) shows the microstructure of the first canning, and 11(B) the second. The results of chemical examination of the welds and brazes and spectrographic analysis of the cladding are given in Table II, showing that the nickel concentration in the second canning was greater than in the first by a factor of two. The correlation between the weld quality of the two canning groups is in agreement with recent work on the effect of nickel content on weld quality.[15]

The most popular welding power supplies available are saturable-reactor-type devices in which the shape of the current wave is directly proportional to the inductance. Obviously, the squarer the waveshape, the higher the cost of the unit because of the additional kilovolt-amperes of magnetic core assembly required.

This study was made to determine, first, the current waveshape that would

Table I. Mating Surface Weld Variation Type IV Current Wave

Piece Designation	Mating Surfaces: At Weld Closure Start	At Weld Closure End	At Weld Mid-Point Start	At Weld Mid-Point End
Male End				
D	II	IV	IV	IV
E	IV	IV	IV	III
F	IV	IV	IV	II
G	IV	III	III	III
H	III	IV	IV	II
I	IV	III	IV	IV
J	IV	IV	III	IV
K	IV	III	IV	II
L	III	IV	IV	III
M	III	IV	IV	IV
Female End				
D	II	III	I	II
E	III	I	III	IV
F	II	I	IV	IV
G	I	III	IV	IV
H	III	II	IV	IV
I	IV			III
J	V			IV
K	IV	I	V	V
L	IV	IV	II	II
M	III	II	IV	I

produce the best alloying in the weld with the fastest welding speed and, second, the open-circuit voltage required to produce good arc starting and a stable arc during welding and crater filling.

Within the limits of available equipment, the current waveshapes were selected to give a uniform distribution of shapes between the two limits from a narrow high-peaked wave with rectification to virtually a square wave, as shown in Fig. 8.

In making the welds for these tests every effort was made to produce all welds with the same dimensions. This is theoretically possible but in practice any slight variation in thickness of the aluminum mass changes its value as a heat sink and thus changes the width of the weld. For a constant value of current, as the mass decreases the weld widens and, conversely, narrows as it increases.

The current programming was arranged to produce a circumferential weld of uniform width for the average fuel element. This is complicated, for reasons not known, by the circumferential weld width variation of consecutive pieces on any one machine. It is not the result of current and arc voltage variation. Recent work indicates the varying width of the completely etched or cleaned area as the cause. If the etched area is not wider than the maximum weld width, the uncleaned area acts as a thermal barrier, limiting the weld width.

Experience has shown that some undesirable conditions occur only once in 100 pieces. Thus, to obtain accurate data on weld quality, this number of pieces must be examined. The pieces must be sectioned diametrically through the closure point on the weld because cracks in the internal structure of the weld occur mainly at this point.

Though many people have done a vast amount of work on the problem of alloying the braze into a homogeneous weld, nothing definitely is known as to why alloying is not complete. Recent work indicates that a constant temperature in the weld will produce virtually complete alloying in the weld. The melting point of the cladding is 655 C (degrees centigrade) and the brazing metal 585 C. There is no detectable difference, within 1 C of the melting point, between the *A-13* braze metal and the *A-13* plus 0.5% nickel.

Previous work shows that it is much more difficult to produce complete alloying of the male weld than for the female. In support of this, the welding rate for the male weld is exactly one third that of the female. Any increase in the welding speed above that used on either the male or female welds decreases their quality.

Table II. Metal Analysis

Component Silicon Nickel, Per Cent

	Canning No. 1		Canning No. 2	
Braze	19.82	0.27	14.81	0.4
Weld bead	0.45	0.58	0.38	1.0

Cladding in both cases medium to strong nickel.

In early quality comparisons between welds produced with types I, II, and III current waveshapes, the weld alloying improved as the current waveshape changed from a type I toward a type III. The improvement in weld quality by the elimination of rectification has been found by others;[16] also, the surface roughness decreases. Fig 12(A) shows the alloying that occurred in one third of the welds produced with the type III current wave, and (B) shows the surface roughness.

In high-speed motion pictures corrugations in the weld surface may be seen forming as the current started to decrease when the electrode was positive. The welds were made at 23½ inches/min (per minute) with argon shielding gas. The type III current wave with exactly the same technique produces satisfactory weld quality. From this it is theorized that the increments of heat per unit time were more uniform as the current was widened and its maximum value decreased. Oscillograms of arc power, current, and voltage for current wave types I, II, and III are shown in Fig. 2. The oscillograms were taken with 200-amp rms in order to make a direct comparison of the power produced in each half-cycle by each type of current wave. In going from an unbalanced type I to a balanced type II current wave, the oscillograms show that the maximum power in a half cycle changes from the half-cycle with the electrode negative to that with the electrode positive. No visible change occurs in the width of the valley between the power pulses, indicating that any gain in weld quality is due to the increased power in the half-cycle in which the electrode is positive, maintaining a more uniform temperature in the weld. There is a large reduction in the surface corrugations of the weld (see reference p. 46).

Comparing the oscillograms of types II and III current waves shown in Fig.

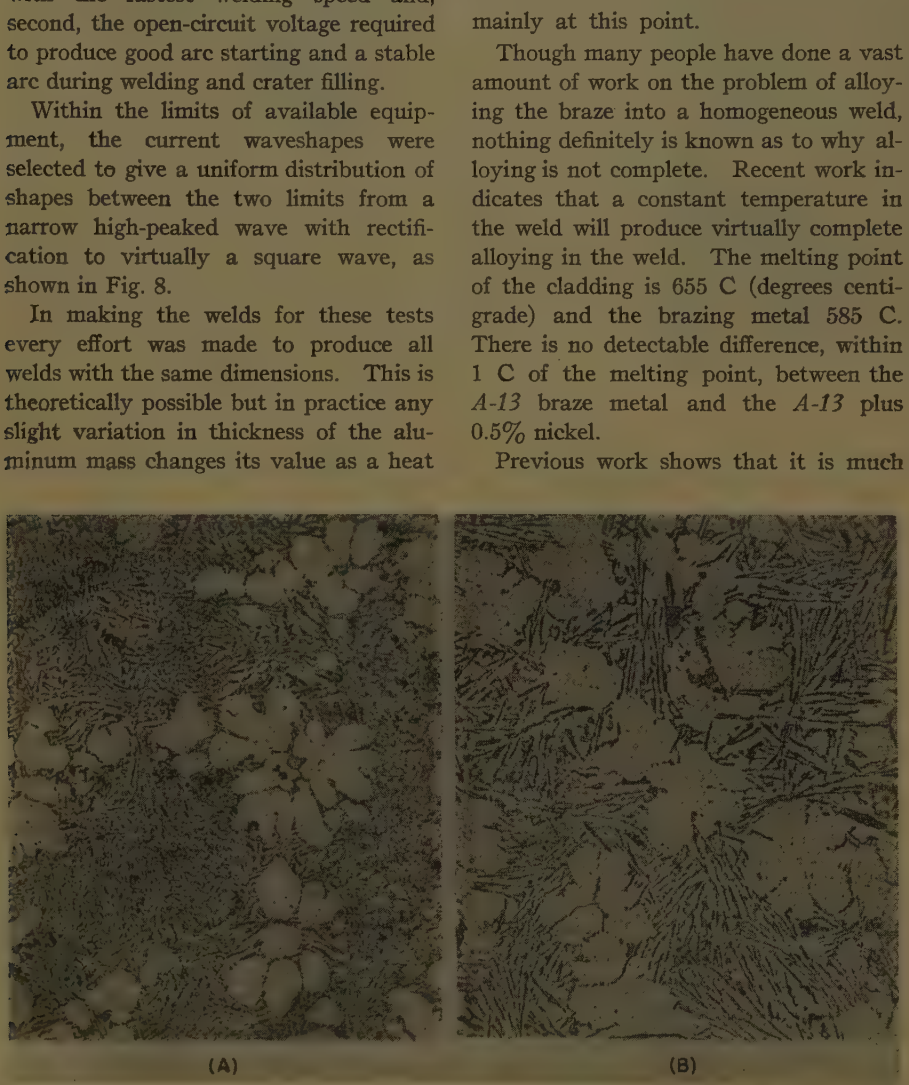
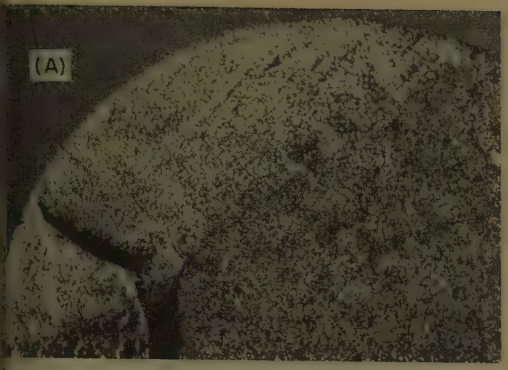


Fig. 11. Braze microstructure, enlarged 250×, 0.5% hydrofluoric acid etch. A—Canning no. 1. B—Canning no. 2



. 12. Weld alloying and weld surface produced type I current wave, 165 amp; argon shielding gas; .3 inches/min. A—Alloying; enlarged 15×. B—Surface; enlarged 2.5×

ows a definite reduction in the width the valley between the power pulses. his is evidenced in the weld quality provement between the two types (see erence 4, p. 53).

Further improvement is shown in the duction of the valley between the power lses of types III and IV current waves own in Fig. 2. Also, the power pulses e developing flat tops, indicating that more uniform temperature is occurring the welds.

One form of square current wave types shown in Fig. 2, types V through VII. comparison of type IV with these ows that in the latter there is a major duction in the width of the valley tween the power pulses. The tops of e power pulses are further broadened d flattened, producing a more uniform mperature in the weld.

Fig. 2, types VIII through XI, gives second form of square wave. A comrison of the previous type of square ve with these shows that the valleys tween the power pulses are identical. here is a difference in the contour of the ps of the two forms.

With current waveshape types I rough XI, a "negative" pulse of power curs as the polarity changes from the ctrode positive to the electrode gative. These pulses of negative wer cannot be accounted for on any sis which include intrinsic arc phemena. From calculations using the own circuit constant of the watt lvanometer and its associated circuit, out one third of the negative power is oduced by the inductance in the galnometer current coil. The remainder probably produced by circuit inductance the watt galvanometer current cirit, and reversal of the charges in the returning energy to the circuit. he difference in amplitudes of the negae power pulses is produced by the difference in those of the restrike potentials.

The effect of the negative power pulses is an apparent narrowing of the width of the power pulses and the widening of the valleys between them. Since the current passing through the watt galvanometer is not the same as that passing through the current galvanometer and since the current galvanometer inductance is negligible, the current trace is the true value. By multiplying the corresponding ordinates of the current and voltage traces, the value of power can be found and it is without evidence of the negative power.[18] The exact alignment of the current and voltage traces can be made by aligning the discontinuity in the current trace with the crest of the restrike voltage when the electrode is positive.

In an attempt to develop a theoretical relationship between the welding current waveshape and its efficiency in alloying in a weld, numerous calculations were made. The values of power were determined graphically from the power oscillogram trace, the instantaneous values of which were known by calibration. This required the use of a function that was independent of the small variations in current, voltage, and power occurring in the arc. The first attempt used form factors of the power pulses, a comparison of which for positive and negative polarities is shown in Table III. The difference in the values is not great enough to produce a significant separation to be used for evaluation. Table IV shows the form factors of the current waveshapes for each cycle of power. Here, there is less separation than in the previous case.

A much better approach was developed by using the ratio of the power in a half-cycle with the electrode positive to that with the electrode negative. These data, in order of improving waveshape are shown in Table V; as the current wave squares up, the ratio increases from type I to type II, then decreases from type II to type IV, and then increase from type V through type XI. Separation and trend are adequate for the evaluation. Change in trend between types I and II is caused by the change from unbalanced to balanced current. The decrease between types II and IV is a very small per cent, and may be caused by errors in the graphical solution. Increasing of squareness in the current wave causes the ratio to decrease between types V and XI. Current wave types IX and X appear to be interchanged by a comparison of the numerical value of the ratio but the trend indicates that they should be in this order, as type X is more square than type IX. Probably this is true because of errors in the graphical method and calculations, as the difference between the two is 0.78%.

Table VI shows results of the half-cycle power ratio and weld evaluation, tabulated according to the ratio of effective power with the electrode positive to effective power with the electrode negative. Good correlation between types III and IV occurs as the weld quality number for both the male and female weld decreases; also, both are from the

Table III. Waveshape Half-Cycle Power Relationships

Current Wave Type	Power, Kw: Average Value	Power, Kw: Effective Value	Form Factor
Positive Electrode			
I	1.798	2.27	1.262
II	3.51	4.68	1.332
III	4.20	4.79	1.141
IV	4.20	4.34	1.033
V	3.51	3.65	1.039
VI	3.22	3.38	1.048
VII	3.38	3.50	1.033
VIII	3.31	3.47	1.047
IX	3.46	3.58	1.035
X	3.45	3.55	1.030
XI	3.43	3.54	1.031
Negative Electrode			
I	4.02	5.17	1.283
II	1.835	2.44	1.330
III	2.20	2.52	1.145
IV	2.07	2.31	1.113
V	1.483	1.500	1.013
VI	1.280	1.310	1.023
VII	1.336	1.358	1.017
VIII	1.400	1.423	1.015
IX	1.375	1.400	1.018
X	1.373	1.398	1.017
XI	1.308	1.340	1.022

same canning group. The larger weld quality number indicates that the male weld is more difficult to produce, in agreement with results of previous work (see reference 4, pp. 28–30). Types IV and V show an increase in weld quality numbers. However, type IV is from the canning group which produced poor weld correlation and type V from that which produced good weld correlation, indicating that the difference is not significant. This is substantiated by a study of Table VII which shows a major improvement in the conditions that are very serious, such as bad welds and cracks. Also, the male weld quality number is higher than that of the female, indicating that it is a poorer weld, which agrees with previous findings. A comparison of types V and VIII indicates that, according to the power pulse ratio, type VIII produces the best weld, while the weld quality number indicates type V. Since both were welded from the same canning group that produces good weld correlation and the numbers are so close together, they will both produce welds of virtually the same quality. In addition, the male weld quality numbers are higher than the female, as expected.

Comparing the power pulse ratios and weld quality numbers for types VIII, IX, and X, shows a definite improvement in quality from type VIII to type X, as shown by the correlation between ascending power pulse ratios and the descending weld quality numbers. Again, the male weld quality numbers are higher than the female, as expected.

The comparison between types X and VI indicates a good correlation between the power ratios but a reverse correlation between the weld quality numbers; this is to be expected, as they are from different canning groups. A study of Tables VII and VIII indicates that for most items type X is a better weld than type VI. However, with previous findings and data it can be concluded from the power pulse ratio correlations that type VI current wave produces a better weld than type X. The male weld quality numbers are higher than the female, as expected.

A comparison of types VI, VII, and XI brings forth some unusual facts: Weld quality numbers indicate that the quality of the welds produced by type VII should be better than that produced by type VI but the power ratios indicate that they should be the same. The data in tables VII and VIII generally indicate that type VII wave weld is better than type VI. Inspection of the power trace of the oscillograms in Fig. 2, types VI and VII reveals a difference in flatness in the tops of the power pulses of the two wave types, indicating that they should not have the same power pulse ratios. In addition, the weld quality number for the male weld is higher than that of the female, indicating proper correlation. Thus it can be stated that in the weld quality produced by the two current wave types, the difference as indicated by the weld quality number is, though measurable, very small.

A comparison of types VII and XI current waveshapes indicates from the power pulse ratios a very good correlation, but from the weld quality numbers, the reverse. The complete weld quality data of Table IX indicates that the total spots and total high silicon quarters agree with the power pulse ratios and with the relationship between the male and female weld quality. Previous data, the power pulse ratios, and knowledge of the materials' weldability lead to the conclusion that a type XI current wave will produce a better weld than a type VII but that the difference in quality cannot be established.

Work in progress in Japan on improved a-c waveshapes indicates that square type current waves increase arc stability and improve welding.[19]

Complete examination of the data points to the conclusion that as an alternating inert-arc welding current wave approaches a square wave, the quality of the fusion weld produced on the Al-bonded aluminum-clad uranium-metal fuel element will improve by a measurable amount.

Table IV. Form Factors of Current Waveshapes

Current Waveshape Type	Power Per Cycle, Kw: Average	Power Per Cycle, Kw: Effective	Form Factor
I	2.81	4.00	1.423
II	2.67	3.73	1.396
III	3.20	3.82	1.182
IV	3.14	4.81	1.53
V	2.50	2.80	1.12
VI	2.25	2.56	1.138
VII	2.36	2.65	1.122
VIII	2.36	2.65	1.122
IX	2.42	2.72	1.122
X	2.41	2.70	1.122
XI	2.37	3.68	1.13

Table V. Current Waveshape Half-Cycle Effective Power Relationship

Current Wave Type	Effective Power, Kw: Electrode Positive	Effective Power, Kw: Electrode Negative	Effective Power, Kw: Total Per Cycle	Ratio Effective Power Electrode Positive to Electrode Negative
I	2.27	5.17	4.00	0.44
II	4.68	2.44	3.73	1.92
III	4.79	2.52	3.82	1.90
IV	4.34	2.31	4.81	1.88
V	3.65	1.50	2.80	2.43
VIII	3.47	1.423	2.65	2.44
IX	3.58	1.40	2.72	2.56
X	3.55	1.398	2.70	2.54
VI	3.38	1.31	2.56	2.58
VII	3.50	1.358	2.65	2.58
XI	3.54	1.34	2.68	2.66

Ratio Relationships

In this section ratios between open-circuit voltage, welding speed, and shielding gas are discussed.

Table VI. Half-Cycle Power Ratio Weld Quality Relationship

Current Wave Type	Ratio Effective Power Electrode Positive to Electrode Negative	Weld Quality Number	Canning Number
I	0.44		
II	1.92		
III	1.90	M 258 F 233	2
IV	1.88	M 230 F 215	2
V	2.43	M 276 F 222	1
VIII	2.44	M 275 F 259	1
IX	2.56	M 245 F 211	1
X	2.54	M 212 F 183	1
VI	2.58	M 265 F 250	2
VII	2.58	M 250 F 259	2
XI	2.66	M 293 F 231	2

M = male weld; F = female weld.

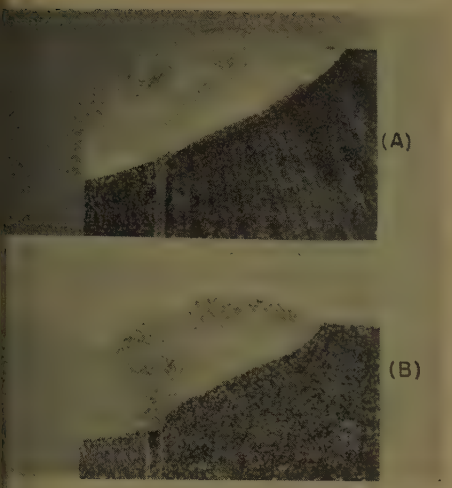


Fig. 13. Weld produced with d-c straight polarity; 52.9 inches/min. A—Argon shielding gas; 150 amp. B—Helium shielding gas; 100 amp, enlarged 10X, caustic etch

Dependency of the arc on open-circuit voltage for stable operation was discovered early in the study of parameters affecting the weld quality of AlSi-bonded aluminum-clad fuel elements (see reference 4, pp. 27-33). For the type III current wave, it was found that 100 open-circuit volts maintained good arc stability under these conditions: argon used as a shielding gas, currents of 140 to 175 amp, 3/16-inch zirconium-tungsten alloy electrodes, a welding speed of 42.4 inches/min, and series current wave balancing capacitors. With a 5/7 argon-helium ratio shielding gas and all other parameters the same, the arc was difficult to start and had a tendency to go out during crater filling.

In evaluating the open-circuit voltage required to produce a stable arc with type XI current wave and a 3/16-inch zirconium-tungsten electrode, it was discovered that an arc that was stable at a welding speed of 17.8 inches/min would go out at a welding speed of 52.9 inches/min. Table X gives results of the tests showing the dependency of a stable arc on welding speed, open-circuit voltage, and shielding gas ratio. Under conditions of 100 open-circuit volts, argon shielding gas, and a welding speed of 52.9 inches/min, the arc was unstable with superimposed high frequency and stable without.

Fig. 2, types I and II, shows the effects of low open-circuit voltage on the third-harmonic distorted sine current waves. The voltage could not rise fast enough and high enough to reignite the arc without a discontinuity in the current. The arc was high-frequency stabilized. Also, as the current nears the end of the pulse, the arc tries to go out and the extinguishing voltage rises, trying to maintain it. The low open-circuit voltage also produces variable height and therefore variable areas of the current pulses. The combination of the variable heights and variable pulse widths produces a continuously varying temperature in the welds.

The time after zero current, in which a glow discharge exists, approaches zero as the open-circuit voltage increases and the current waveshape becomes square. When the electrode becomes positive, the time in which the glow discharge exists approaches zero as the waveshape becomes square; this is shown in Fig. 8 by the discontinuities in some of the current traces as the current rises above zero. The discontinuities are the result of too few electrons and possibly positive ions available at that instant to support a true arc. During the period of transition, enough heat is added by particle acceleration from the rising voltage to produce sufficient electron emission and to permit a true arc to form. When the refractive tungsten electrode becomes negative, enough electrons are available to form essentially a continuous true arc; see Fig. 8. No discontinuities are recorded as the cur-

Table VII. Individual Weld Quality

Lot No.	Individual Weld Section Quality Number of Each Class Per Lot: Class I, Excellent	Class II, Good	Class III, Fair	Class IV, Poor	Class V, Bad	Class VI, Internal Cracks
1M	12	89	52	226	25	15
1F	49	101	73	164	34	7
2M	36	101	111	160	3	4
2F	94	106	93	112	4	9
3M	70	116	86	105	16	
3F	140	125	83	45	19	1
4M	18	52	114	214	12	4
4F	84	74	64	150	17	1
5M	4	45	74	213	31	48
5F	76	54	61	179	33	10
6M	10	45	29	207	29	79
6F	53	49	58	188	44	16
7M	5	68	58	185	19	22
7F	108	75	55	111	42	17
8M	6	106	73	190	19	11
8F	56	135	43	151	21	3
0M	10	36	25	316	9	4
0F	103	80	48	128	10	31

Table VIII. Individual Weld Quality Tabulation

Lot No.	No. of Surface Penetrations by Silicon, Spots per Weld: 1	2	3	4	5	6	7 or More	Generally High Silicon in Quarter Circumferences of the Weld, Quarters: 1st	2nd	3rd	4th
1M	6	7	2								
1F	7	7	3	4	1		1				
2M	8	6	3	1							
2F	2	4	7		2	1					
3M	11	6	6	1							
3F	4	2	4	2	1						
4M	17	14	12	8	1	1					
4F			2	1		1	1				
5M	4	13	13	11	14	17	49	3	15	3	
5F	4	7		8	3	2	6	8	8	3	1
6M	11	16	10	15	9	5	4	16	25	4	2
6F	6	2	4		1		1	31	13	1	1
7M	1	1	12	11	5	9	5	19	30	7	8
7F	3	5	5	4	6	9	3	12	12	4	6
8M	1	1	1	3			3	22	30	13	17
8F			1	1				22	25	4	1
0M	7	7	6	2	4			23	15	1	7
0F	7	2	2	1				19	3		

Table IX. Complete Weld Quality Tabulation

Lot No.	Pieces in Lot	Waveshape Types	Weld Quality Number	Total Spots	Total High Silicon Quarters
8M	104	III	258	18	189
8F			233	7	88
7M	103	IV	230	197	132
7F			215	149	72
4M	102	V	276	87	
4F			222	23	
1M	105	VIII	275	26	
1F			259	58	
2M	105	IX	245	33	
2F			211	47	
3M	107	X	212	49	
3F			183	39	
5M	106	VI	265	546	42
5F			250	123	37
6M	101	VII	250	237	86
6F			259	34	64
0M	102	XI	293	67	84
0F			231	21	25

M = male weld; F = female weld.

rent increases below the zero axis. From theoretical considerations there must always be a finite time, after the voltage starts to increase from zero, in which a glow discharge exists. That a glow discharge does exist and that there is an electron deficiency, (though not shown in the current trace) is shown by the peaks on the voltage traces after the point of zero current; see Fig. 2, types IV through IX.

With increasing open-circuit voltage and squaring up of the current wave, the extinguishing voltage required to maintain the arc at the end of the pulse is at least as great as the burning voltage; see the round-cornered ends of voltage traces in Fig. 2, types IV through XI. The extinction voltage is shown as a definite spike indicating insufficient voltage available to maintain a true arc; see Fig. 2, Types I, II, and III.

The results of the tests indicate that an open-circuit potential of 130 volts would produce the required arc stability with all the gas ratios tested. This value was used throughout the tests on types V through XI, current waveshapes producing excellent arc stability with a 5/7 argon-helium shielding gas ratio. This is because the maximum restrike voltage occurs at virtually zero current.

The maximum open-circuit potentials allowed by equipment limitations were 80 volts for types I and II and 100 volts for type III current waveshapes. Type IV had a constant open-circuit potential of 165 volts.

The use of a high open-circuit voltage to maintain a stable arc has the following advantages:

1. Minimum radio-frequency interference problems when high frequency is used for arc starting only.
2. High-frequency stabilization tends to produce a weld with a rougher margin and surface.
3. The arc is more stable, having less tendency to wander, producing a smoother weld.
4. Increased welding speed.
5. The arc starts more quickly and smoothly.

Table X. Type XI Current Wave, Arc Stability Relationships for Variable Gas Ratios

Open Circuit, Volts	Shielding Gas Volume Ratio		High Frequency Stabilization		Arc Stability, Inches/Min							
					17.8				52.9			
	Argon	Helium	Yes	No	Unstable	Almost Stable	Stable	Very Stable	Unstable	Almost Stable	Stable	Very Stable
80	x			x				x	x			
	x		x		x				x			
	5	7	x		x							
90	x			x					x			
	5	7	x		x							
95	x			x						x		
	5	7	x				x					
100	x		x							x		
	x			x							x	
	5	7	x				x					
105	x			x								
	5	7	x				x					
	5	7		x	x							
110	5	7		x			x					
	3	9		x	x							
120	5	7		x						x		
	3	9		x					x			
125	5	7		x							x	
	3	9		x					x			
130	5	7		x								x
	3	9		x								x

Conclusions

From test results and demonstrations the following conclusions to may be established.

Improvement of fusion weld alloying with square-type current waves in the closure weld of AlSi-bonded aluminum-clad uranium-metal fuel elements is established. Alloying improves as the current waveshape changes from an asymmetrical 180-degree out-of-phase third-harmonic distorted sine wave toward a symmetrical square wave.

A positive agreement exists between the ratio of the power pulse with the electrode positive to the power pulse with the electrode negative and the weld quality number, showing that alloying in the weld improves as the current waveshape changes toward a square wave. Also weld alloying can be measured by a number.

Weld alloying improves as the temperature in the weld is held more constant and is concentrated by a more constant value of current in each pulse. This is supported by work in progress with d-c straight polarity using argon and helium shielding gases.[20] With d-c straight polarity the electrode is small enough that the arc diameter decreases and a plasma jet forms, increasing agitation of the molten metal and weld penetration. Sections of representative welds from a lot of ten pieces are shown in Fig. 13. The weld in Fig. 13(A) was produced with argon and the one in (B) with helium. The argon weld is class II, and the helium weld class I. There was no evidence to indicate that the weld quality throughout a lot for either gas would be less than shown.

A relationship exists between welding speed, shielding gas ratios, and the open-circuit; voltage required to produce

stable arc without the use of high-frequency stabilization and also the use of superimposed high-frequency current will contribute to an unstable arc condition with open-circuit voltages that will produce a stable arc.

As the current wave becomes square, the welding speed may be increased without adversely affecting weld quality.

In changing the shielding gas mixture from a pure argon toward a pure helium the open-circuit voltage required to maintain arc stability with and without superimposed high-frequency current increases.

From the test results it may be stated that the more desirable characteristics for a welding power supply to fusion weld the closure of AlSi-bonded aluminum-clad uranium-metal fuel elements with the a-c tungsten inert-gas-shielded arc are:

1. A balanced current waveshape that is as good as type XI, Fig. 8.
2. Open-circuit potential of 130 volts for argon-helium mixtures of up to 3 argon, 9 helium by volume.

References

1. Arc Characteristics and Their Significance in Welding, **D. R. Milner, G. R. Salter, J. B. Wilkinson.** *British Welding Journal*, London, England, vol. 7, no. 2, Feb. 1960, pp. 73–88.

2. Metal Transfer in Inert-Gas Shielded-Arc Welding, **J. C. Needham, C. J. Cooksley, D. R. Milner.** *Ibid.*, p. 107.

3. Gaseous Conductors (book), **J. D. Cobine.** Dover Publications, Inc., New York, N. Y., 1958, pp. 302, 343.

4. Fusion Welding of AlSi Bonded Fuel Elements, **Thomas B. Correy.** *AEC Research and Development Report no. 48978*, General Electric Company, Richland, Wash., Mar. 11, 1957, p. 15.

5. Gas-Shielded Arc Cleaning, **Ontario H. Nestor.** *U. S. Patent no. 2,906,857*, Sept. 29, 1959.

6. Current Density of the Arc Cathode Spot, **J. D. Cobine, C. J. Gallagher.** *The Physical Review*, New York, N. Y., vol. 74, no. 10, Nov. 15, 1948, pp. 1524–30.

7. Plasma-Energy Transfer in Gas-Shielded Welding Arcs, **H. C. Ludwig.** *Welding Journal*, New York, N. Y., vol. 38, no. 7, July 1959, p. 298–S.

8. The Calorimetric Study of the Arc, **P. P. Alexander.** *AIEE Transactions*, vol. 49, Apr. 1930, pp. 519–23.

9. Arc Welding, **A. U. Welch, Jr.** *U. S. Patent no. 2,472,323*, June 7, 1949.

10. Heliarc Welding of Aluminum, **Thomas B. Correy.** *AEC Paper no. HWSA-1731*, General Electric Company, Richland, Wash., Mar. 17–18, 1959, p. 22.

11. Extinction of an A-C Arc, **J. Slepian.** *AIEE Transactions*, vol. 47, Oct. 1928, pp. 1398–1408.

12. The Arc in Circuit Interrupters, **J. Slepian.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 214, Oct. 1932, p. 413.

13. Probe Measurements and Potential Distribution in Copper A-C Arcs, **W. G. Dow, S. S. Attwood, G. S. Timoshenko.** *AIEE Transactions*, vol. 52, Sept. 1933, pp. 926–33.

14. Reignition of Metallic A-C Arcs in Air, **S. S. Attwood, W. G. Dow, W. Krausnick.** *Ibid.*, vol. 50, Sept. 1931, pp. 854–70.

15. The Influence of the Nickel Content of the Brazing Alloy on Fuel Element Weld Quality, **G. R. Hanson.** *AEC Confidential Report no. HW-63138* (classified), General Electric Company, Richland, Wash., Dec. 21, 1959.

16. The Effect of D-C Component in A-C Inert-Gas Welding of Aluminum, **G. J. Gibson, G. R. Rothschild.** *Welding Journal*, vol. 27, Oct. 1948, Research Supplement, pp. 4965–5015.

17. The 'Twin-Argon' Welding Process, **J. A. Donelan.** *British Welding Journal*, vol. 4, no. 1, Sept., 1954, pp. 403–08.

18. Development of an Improved Current Wave Shape Alternating Current Welding Power Source, **Thomas B. Correy.** *AEC Report no. HWSA-1730*, General Electric Company, Richland, Wash., Oct. 13, 1959, p. 3.

19. ETL Type Arc Welder, **E. Sugihara, S. Kikuchi, K. Yada.** *Japan Welding Society Journal*, Tokyo, Japan, vol. 28, no. 10, Oct., 1959, pp. 31–36. A translation is in the library of the Hanford Works of the U. S. Atomic Energy Commission operated by General Electric Company, Richland, Wash.

20. Welding Fuel Slugs D-C Helium Process, **J. E. Harpring, J. W. Kelker, Jr.** *AEC Report no. DPSP 59-25-26, Met-N21-185* (classified), E. I. du Pont de Nemours and Company, Atomic Energy Division, Savannah River Plant, Savannah, Ga., Oct. 1959, pp. 2–23.

Experimental Determination of the Frequency Response of a Linear Transfer Function for Arbitrary Transient Inputs of Finite Duration

W. W. WIERWILLE
NONMEMBER AIEE

IN MANY PRACTICAL situations when frequency-domain information about a process is required, the input is not sinusoidal and of a constant-frequency steady-state nature. For example, it may be necessary to obtain the frequency response of the changing linear plant in an adaptive control system at a specified time. A problem arises in that the plant is controlling some quantity and usually cannot be disturbed by measurement devices. Consequently, the measurement must be made using only input and output information of the linear plant, which is not usually sinusoidal in form. The problem then is one of obtaining frequency-domain information about the system from the given input.[1]

This paper outlines a simple way of determining complete frequency-domain information when the input is any arbitrary transient of finite duration. The method lends itself to incorporation in physical equipment and can be rigorously verified in theory.

Although its chief value lies in the application previously stated, the method has two additional valuable uses. First, it can be used to convert an impulse, step, or ramp response in the time domain to the frequency response of a system in the frequency domain. Thus, given the impulse response of a control system, the frequency response is obtainable. Second, the method can be used to obtain the Fourier transform of an arbitrary waveform of finite-time duration.

The bulk of prior work in this area has been analytical in nature.[2] The frequency-domain information usually has been gained only by laborious calculations, often requiring digital computers for solution. There appears to be a genuine need for a device which can quickly compute, in at least a semiautomatic manner, the necessary frequency-domain information. The experiment described herein is successful in quickly obtaining this information when the inputs are of a class which included those made up of an arbitrary transient of finite duration, followed by an interval in which the input was fixed at some value.

Theoretical Basis

Usually, when the frequency-domain information of a control system is required, a sinusoidal generator of variable frequency is connected to the input, as shown in Fig. 1. After the transient resulting from the change in input has decayed to zero, the amplitude and phase

Paper **61-745**, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Ithaca, N. Y., June 18–23, 1961. Manuscript submitted February 23, 1961; made available for printing May 8, 1961.

W. W. Wierwille is with Cornell University, Ithaca, N. Y.

The author wishes to thank Professor W. E. Meserve of Cornell University for his advice and assistance.

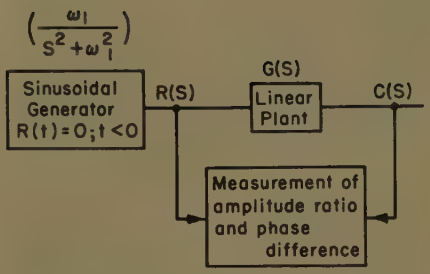


Fig. 1. Frequency-response test procedure for sinusoidal input

of the control system's output are compared, respectively, with those of the input. From this information, a point on both the amplitude and phase plots can be determined. For this arrangement, the output $C(t)$ of the control system, when a sinusoidal generator is connected at $t=0$, has the Laplace transform

$$C(s)=\frac{\omega_1}{s^2+\omega_1^2}\,G(s) \tag{1}$$

where $G(s)$ is the complex transfer function of the control system, and ω_1 is the frequency of the measurement.

A band-pass filter has a related transform for its transfer function. Such a filter could be made to have the transfer function given by

$$\frac{\theta_0(s)}{\theta_i(s)}=\frac{\omega_1 s}{s^2+2\sigma\omega_1 s+\omega_1^2} \tag{2}$$

If the filter had an infinite Q, then σ would be zero and the transfer function would become

$$\frac{\theta_0(s)}{\theta_i(s)}=\frac{\omega_1 s}{s^2+\omega_1^2} \tag{3}$$

Two of these filters can be used to obtain the frequency response of the system if they are employed in a test arrangement where the input is a step as shown in Fig. 2.

Proof that the frequency response is obtainable, when the input is a step and ideal filters are used, lies in the fact that the transform of the output $C_1'(s)$ is identical to that of the transform $C(s)$, as given in equation 1 for a sinusoidal test:

$$C_1'(s)=\left(\frac{1}{s}\right)\left(\frac{\omega_1 s}{s^2+\omega_1^2}\right)G(s)$$

Similarly, the input functions are equal:

$$R_1'(s)=R(s)=\left(\frac{1}{s}\right)\left(\frac{\omega_1 s}{s^2+\omega_1^2}\right) \tag{4}$$

Therefore, if the filters are ideal, the transfer function of any linear plant $G(s)$ can be obtained from a step input by waiting until the system transient has decayed to zero and then making measurements of amplitude ratio and phase difference at the outputs of the two filters.

An example of measuring a simple second-order transfer function with a step input will help to clarify the procedure. The Laplace transform of the response of the second filter is

$$C_1'(s)=\left(\frac{1}{s}\right)\left(\frac{1}{1+\frac{2\sigma s}{\omega_2}+\frac{s^2}{\omega_2^2}}\right)\left(\frac{\omega_1 s}{s^2+\omega_1^2}\right)$$

where

$$G(s)=\left(\frac{1}{1+\frac{2\sigma s}{\omega_2}+\frac{s^2}{\omega_2^2}}\right)$$

The time response is

$$C_1'(t)=\left[\frac{\omega_1\omega_2^2}{[(\omega_2^2-\omega_1^2)^2+4\sigma^2\omega_1^2\omega_2^2]^{1/2}}\right]\times$$

$$\left[\frac{1}{\omega_2\sqrt{1-\sigma^2}}\,e^{-\sigma\omega_2 t}\sin\left(\omega_2\sqrt{1-\sigma^2}\,t-\Omega_2\right)+\frac{1}{\omega_1}\sin\left(\omega_1 t-\Omega_1\right)\right]$$

where

$$\Omega_1=\tan^{-1}\frac{2\sigma\omega_1\omega_2}{\omega_2^2-\omega_1^2}$$

and

$$\Omega_2=\tan^{-1}\frac{-2\sigma\omega_2^2\sqrt{1-\sigma^2}}{\omega_1^2-\omega_2^2(1-2\sigma^2)}$$

After a time, equal in value to five time constants of the exponential function, the transient is negligible, so that the response is given by

$$C_1'(t)\Big|_{t\geq\frac{5}{\sigma\omega_2}}=\frac{\omega_2^2}{[(\omega_2^2-\omega_1^2)^2+4\sigma^2\omega_1^2\omega_2^2]^{1/2}}\times\sin\left(\omega_1 t-\Omega_1\right)$$

The output of the first filter has the transform given by equation 4, which is a sinusoid:

$$R_1'(t)=\sin\omega_1 t$$

The ratio of the amplitude of the second filter to that of the first filter is

$$\frac{\omega_2^2}{[(\omega_2^2-\omega_1^2)^2+4\sigma^2\omega_1^2\omega_2^2]^{1/2}}$$

which is exactly the magnitude of the transfer function obtained by substituting $j\omega_1$ for s in the transfer function $G(s)$. In addition, the phase difference is Ω_1, which is identical to the complex phase angle of the transfer function at ω_1.

A question might be raised at this point as to whether or not this is simply a trick in block-diagram algebra. The answer is that it is not. It happens that a sinusoidal generator output and an ideal filter transfer characteristic have similar Laplace transforms and use is made of this fact. Also, it is certainly true that the system pictured in Fig. 1 is very different from that of Fig. 2.

The foregoing general results can be extended easily to the case of an impulse or ramp input by utilizing appropriate differentiation or integration as shown in Fig. 3.

The transient which occurs in the output of the second filter is identical to that occurring in the output of the system shown in Fig. 1, where the input is

$$R(t)=\begin{cases}\sin\omega_1 t; & t\geq 0\\ 0; & t<0\end{cases} \tag{5}$$

The significance of this fact is that the frequency-domain information is being obtained as quickly as would be possible if a sinusoidal generator were connected

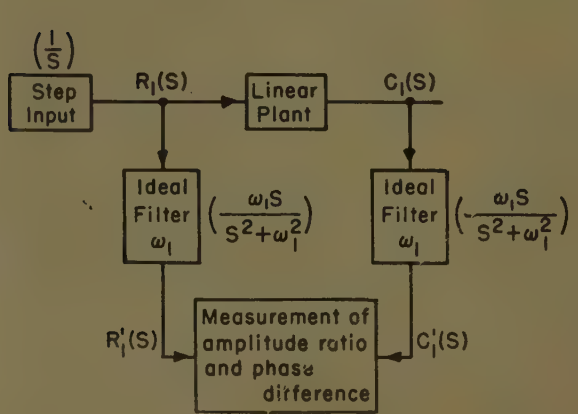


Fig. 2 (left). Frequency response test procedure for step inputs

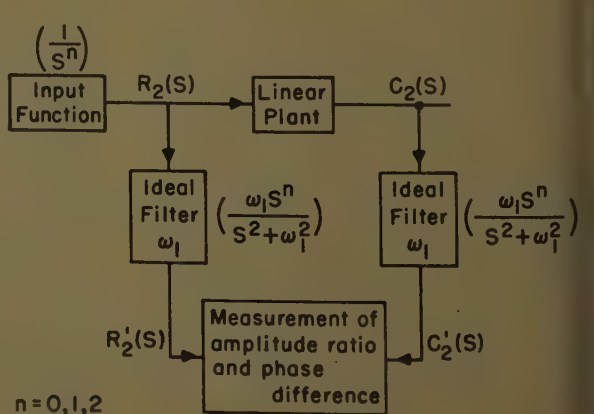


Fig. 3 (right). Frequency - response test procedure for impulse, step, or ramp inputs

to the input and measurements were taken immediately following the transient.

Extension of Theoretical Results

The previous results, which are useful for a group of problems, also can be extended to obtain the frequency-domain characteristic of a linear transfer function when the input is an arbitrary transient of finite duration.

Let the input function to the control system under test be defined as

$$R_3(t)=\begin{cases}A & ;\ t<0\\ R_{3_1}(t); & 0\le t\le b\\ B & ;\ t>b\end{cases} \tag{6}$$

where A and B are finite constants, and

$R_{3_1}(t)$

is a variable which remains finite. For the purpose of proving the extension of results, an ideal filter is placed between the input function and the control system as shown in Fig. 4. If it can be shown that $R_3'(t)$ eventually becomes sinusoidal with this configuration, then it is possible to obtain the frequency response of $G(s)$. The procedure used in the figure is of no practical value and is used solely as a step in the proof. (Since the input has value prior to $t=0$, a Fourier transform analysis is used.)

The input to the control system is

$$R_3'(j\omega)=R_3(j\omega)\frac{j\omega\omega_1}{\omega_1^2-\omega^2}$$

For the convolution of two time functions, it is known that

$$\int_{-\infty}^{\infty}F_a(t-\xi)F_b(\xi)d\xi = \mathfrak{F}^{-1}[(F_a(j\omega))(F_b(j\omega))] \tag{7}$$

where $F_a(t)$ and $F_b(t)$ are the two time functions, and $\mathfrak{F}^{-1}[\ ]$ indicates the inverse Fourier transform.[3] Let

$$F_a(j\omega)=\frac{\omega_1}{\omega_1^2-\omega^2}$$

and

$$F_b(j\omega)=j\omega(R_3(j\omega))$$

then

$$F_a(t)=\sin\omega_1 t$$

$$F_b(t)=\frac{d}{dt}[R_3(t)]+C_0=R_3^{(1)}(t)$$

and

$$(F_a(j\omega))(F_b(j\omega))=R_3'(j\omega)$$

The time function of the input can be written

$$R_3'(t)=\int_{-\infty}^{\infty}\sin\omega_1(t-\xi)R_3^{(1)}(\xi)d\xi \tag{8}$$

where

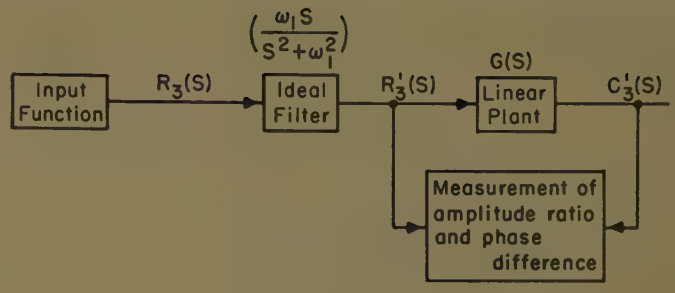


Fig. 4. Conversion of transient input into sinusoidal input for measurement of frequency response

$$R_3^{(1)}(t)=\begin{cases}R_{3_1}^{(1)}(t); & 0\le t\le b\\ 0 & ;\ \text{elsewhere}\end{cases}$$

and $R_{3_1}^{(1)}(t)$ includes the finite impulses that may occur at $t=0$ and at $t=b$.

Under the condition that $t>b$, the system input is

$$\begin{aligned}R_3'(t)\big|_{t>b}&=\int_0^b\sin\omega_1(t-\xi)R_3^{(1)}(\xi)d\xi\\ &=(\sin\omega_1 t)\int_0^b(\cos\omega_1\xi)R_3^{(1)}(\xi)d\xi-\\ &\qquad(\cos\omega_1 t)\int_0^b(\sin\omega_1\xi)R_3^{(1)}(\xi)d\xi\\ &=C_1\sin\omega_1 t+C_2\cos\omega_1 t\\ &=C\sin(\omega_1 t+\phi)\end{aligned} \tag{9}$$

where C and ϕ are constants.

A second proof can be obtained by considering the energy in the filter after $t>b$.

It is thus shown that $R_3'(t)$ for $t>b$ is a sinusoidal function with frequency ω_1 and with constant amplitude and phase. Accordingly, the frequency response of $G(s)$ can be obtained.

The measurement technique of Fig. 4 is, of course, unsuitable because the original input information to the system is modified by the filter. A system which does not modify the input information but allows the equivalent measurement of the transfer function is shown in Fig. 5. The relationships existing for the configuration of Figs. 4 and 5 are shown in the next statement: If $R_3(s)=R_4(s)$, then $R_3'(s)=R_4'(s)$ and $C_3'(s)=C_4'(s)$. As a result of these relationships, when the constant C of equation 9 does not equal zero, the frequency response of $G(s)$ at ω_1 can be obtained.

The output of the first filter $R_4'(t)$ may have zero amplitude for $t>b$; that is, $C=0$. This condition occurs when the input $R_4(t)$ has no energy at the frequency of measurement ω_1. Since $R_4(t)$ is finite, there can be no interval of frequency over which there is no energy. However, isolated points on the frequency axis can have zero amplitude. These points are insignificant because their number is small, and slight detuning of the filters produces a frequency at which the amplitude is not zero.

The positions of these points can be found in two ways. Analytically, the method is to take the Fourier transform of the input function and solve for the zeros of this transform. Experimentally, the positions are determined by the fact that no energy appears in the outputs of the two filters at these points.

From a knowledge of Fourier transforms, it is possible to show that the greater the spread of a function in the time domain, the narrower is its corresponding frequency domain function. For this reason, the input function should be kept within a reasonably small interval (determined by the dynamic range of the measuring equipment) in order to insure sufficient energy over the spectrum of measurement. In addition, the length of the input transient in time adds directly to the computing time. Therefore, if the transient is long, computation is slowed. A second consequence of long transient inputs is the higher accuracy required in matching the tuning of the two filters.

Now it is clear that by using ideal filters, the frequency response of any linear plant $G(s)$ can be obtained, except at a finite number of distinct points, from a transient input of arbitrary form and finite duration. By waiting until the input and system transients have disappeared and, subsequently, making measurements of amplitude ratio and phase difference, the frequency response is determined.

This approach to system transfer-function determination can also be applied in determining the Fourier transform of arbitrary time functions of finite duration. Fig. 6 shows a possible configuration for such determinations. The triggering device produces a unit step at the start of the transient and holds the step until the measurement is completed. In this way, the arbitrary transient is made to appear as a transfer function, and it is handled by the first method described.

Physical Realization

The degree to which the previous theoretical results can be used in practice is dependent entirely upon the ability to synthesize accurately the transfer func-

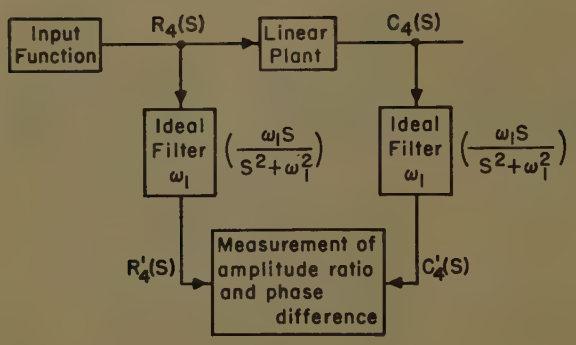


Fig. 5. Frequency-response test procedure for arbitrary finite-duration transient inputs

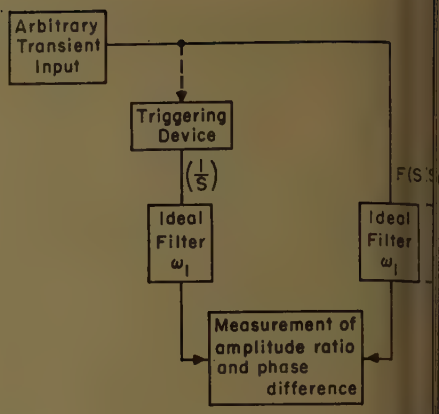


Fig. 6. Test procedure for obtaining Fourier transform of arbitrary finite-duration transient

tions of the ideal filters. The most direct approach is to attempt synthesis of these functions by using analog computer components, which are readily available. In many instances, it may be desirable to program a small analog computer temporarily to produce the required filters. Since direct-voltage drift is of no consequence, the amplifiers may be quite inexpensive without causing loss of accuracy. The equation of the desired filter is

$$\frac{\theta_0(s)}{\theta_i(s)}=\frac{\omega_1 s}{s^2+\omega_1^2}$$

and the corresponding equation in differential operators is

$$\ddot{\theta}_0-\omega_1\dot{\theta}_i=-\omega_1^2\theta_0 \qquad (10)$$

This can be synthesized exactly, using the computer diagram of Fig. 7. Any inaccuracy is caused solely by inaccuracy of the computer components. If it occurs in the potentiometer settings, capacitor values, or resistor values, it may be compensated for; hence, the essential inaccuracies will result from the nonideal nature of the amplifiers themselves.[4] Accordingly, a brief study of these amplifiers will be undertaken.

Operational amplifiers are of two general types, the first of which achieves high gain by means of controlled positive feedback. This is accomplished by applying the output of a high-gain stage to a cathode follower, which in turn is used to drive the cathode of a tube in the high-gain stage. Accordingly, positive feedback is incorporated, and the gain becomes exceedingly high at low frequencies. The essential block diagram of such an amplifier is shown in Fig. 8. The pole at ω_c is the result of the shunt capacitance at the output of the high-gain stage.

The gain of the amplifier at low frequencies in the absence of positive feedback is

$$\left(\frac{K_F}{\omega_c}\right)$$

The positive feedback gain is β. Therefore, the closed-loop transfer function is

$$\frac{E_0(s)}{E_g(s)}=\frac{-K_F}{s+\omega_c-\beta K_F}=\frac{-\dfrac{K_F}{\omega_c-\beta K_F}}{1+\dfrac{s}{\omega_c-\beta K_F}}$$

The transfer function has a low-frequency gain equal to

$$\frac{-K_F}{\omega_c-\beta K_F}$$

and a pole at the frequency $\omega_c-\beta K_F$.

The second type of operational amplifier uses cascaded stages of gain to achieve the high gain required. In one typical case, the first stage has a pole near the origin, followed by a zero at a higher frequency. The second stage has a pole at that higher frequency. The resulting transfer function is

$$\frac{E_0(s)}{E_g(s)}=K_1\left(\frac{1+s/\omega_{\text{high}}}{1+s/\omega_{\text{low}}}\right)\times K_2\left(\frac{1}{1+s/\omega_{\text{high}}}\right)=\frac{K_1K_2}{1+s/\omega_{\text{low}}}$$

Therefore, to a good approximation, operational amplifiers in general have a transfer function at low frequencies, characterized by

$$\frac{E_0(s)}{E_g(s)}=\frac{-K}{1+s/\omega_e} \qquad (11)$$

If an operational amplifier is used as a summing amplifier, the input grid voltage $E_g(s)$ is governed by the equation:

$$E_g(s)=E_i(s)\frac{R_f}{R_i+R_f}+E_0(s)\frac{R_i}{R_i+R_f}$$

where $E_i(s)$ is the summing-amplifier input. The input grid-to-ground capacitance is neglected in this equation because its effect is small at low frequencies. In addition, this small effect can be made negligible by using nominal input resistance of 50 kilohms instead of 1 megohm.

The over-all transfer function for a summing amplifier is

$$\left.\frac{E_0(s)}{E_i(s)}\right|_{\text{sum}}=\frac{-R_f}{R_i+\dfrac{(R_i+R_f)(1+s/\omega_e)}{K}}$$

By substituting $1/Cs$ for R_f, the equation for an integrator is obtained

$$\left.\frac{E_0(s)}{E_i(s)}\right|_{\text{int}}=\frac{-1}{R_iCs+\dfrac{(1+R_iCs)(1+s/\omega_e)}{K}}$$

The characteristic equation of the system of these amplifiers, connected as in Fig. 7, is

$$\left[\frac{E_0(s)}{E_i(s)}\right]^2\left|_{\text{int}}\right.\left.\frac{E_0(s)}{E_i(s)}\right|_{\text{sum}}-1=0$$

where it is assumed that the input resistors to the amplifier are determining the measurement frequency. If higher-order effects are neglected (those involving $1/K^2$ and $1/K^3$) and if $\omega_1=1/R_iC$, $R_i=R_f$, then the characteristic equation becomes

$$1+s\left(\frac{2}{K\omega_1}\right)+s^2\left(\frac{1}{\omega_1^2}+\frac{2}{K\omega_1\omega_e}+\frac{4}{K\omega_1^2}\right)+s^3\left(\frac{4}{K\omega_1^2\omega_e}\right)=0 \qquad (12)$$

Under the assumption that K is large, this equation can be factored into the form

$$\left[1+s\left(\frac{2}{K\omega_1}-\frac{4}{K\omega_e}\right)+s^2\left(\frac{1}{\omega_1^2}+\frac{2}{K\omega_1\omega_e}+\frac{4}{K\omega_1^2}\right)\right]\times\left[1+s\left(\frac{4}{K\omega_e}\right)\right]=0 \qquad (13)$$

whereas the desired characteristic equation is

$$1+s^2/\omega_1^2=0 \qquad (14)$$

In order that equation 13 may approach the desired characteristic equation, both the internal gain K and the gain-bandwidth product $K\omega_e$ of the operational amplifiers must be large. Most standard commercial operational amplifiers have a gain in excess of 10^4. This gain is sufficient at low measurement frequencies [below 250 rad/sec (radians per

econd)] to remove those terms depend-nt solely upon K and ω_1. In addition, everal of these amplifiers have a sufficient ain-bandwidth product to make the re-aining error terms small enough for recision measurement.

In the case where the operational ampli-er does not have a sufficient gain-band-idth product, simple compensation can e incorporated in the connection of the hree amplifiers representing the filter. ig. 9 shows the computer diagram which s modified to include this compensation. Operational amplifier no. 3 has a com-ensating resistor in series with the feed-ack capacitor. The value of this re-istor is set to compensate for the root f the characteristic equation at $s=-K\omega_e/4$. This resistor value is constant nd independent of the measuring fre-uency. After this compensation, the haracteristic equation is

$$1+s\left(\frac{2}{K\omega_1}-\frac{4}{K\omega_e}\right)+s^2\left(\frac{1}{\omega_1^2}+\frac{2}{K\omega_1\omega_e}+\frac{4}{K\omega_1^2}\right)\Bigg]=0$$

Hence, the damping term is

$$=\left(\frac{1}{K}-\frac{2\omega_1}{K\omega_e}\right)$$

This damping term is usually negative nce $\omega_e<2\omega_1$ and is of sufficient magni-ude to require compensation. The mount of compensation required is a near function of the measuring fre-uency, and it is not necessary to com-ensate for the first term of the damping ecause it is negligible. In order to avoid he use of a third frequency-adjustment ontrol, a special positive damping com-ensation is used on operational amplifier o. 2. The 200-ohm resistor and R_2 form voltage adder, which is ahead of the fre-uency-adjusting resistance of the stage. y using this method of compensation, he value of R_2 is fixed for all frequencies at

$$_2\approx K\omega_e\times 100$$

he exact value of this resistance should e determined experimentally because utput impedance of the summing ampli-er, slight loading of input resistor, and omponent inaccuracies may cause error.

The only serious error remaining, although it is small and has a stable value, is that of measurement frequency.

Therefore, it can be compensated for simply by calibrating the frequency-adjustment resistances to the corrected value of frequency. In other words, the loop gain of the 3-amplifier circuit is increased slightly.

The means of compensation for the unwanted root, the damping, and the measurement frequency is one which produces a high-precision filter. Slight changes in internal parameters will not cause appreciable errors. The compensation is simple and inexpensive, and it should be effective to approximately 100 cps (cycles per second).

If the operational amplifiers are being designed for use as components of the filter, the necessary gain and gain-bandwidth product can be achieved by using controlled positive feedback. The bandwidth of a high-gain stage can be traded for low-frequency gain. For example, a pentode which has a gain of 40 db (decibels) over a bandwidth of 100 kc can be converted into a stage with a gain of 100 db and a bandwidth of 100 cps. The relation between the original response a and the response with controlled positive feedback b is pictured in the plot of Fig. 10. The gain-bandwidth product is constant, regardless of the degree of positive feedback. The gain and the gain-bandwidth product of b are sufficient to make all errors in the characteristic equation negligible.

The need for compensation or special design arises only when high accuracy or large measurement bandwidth is required Hence, their use is not required in most cases.

Sample Analysis of Errors

To determine the magnitude of errors created as a result of nonideal filters, the test of a second-order linear plant will be analyzed, using filters with damping other than zero. Error analysis for a general transfer function $G(s)$ is difficult; therefore, it becomes necessary to examine a specific system. For purposes of examination, a system with the test configuration of Fig. 2 will be used, except that the equivalent transfer function of the filters will be

$$\frac{\theta_0(s)}{\theta_i(s)}=\frac{\omega_a s}{s^2+2\sigma_a\omega_a s+\omega_a^2} \qquad (15)$$

In this case, the transfer function of the 3-amplifier filter is being approximated by a second-order system having a resonant frequency ω_a and a damping σ_a. The quantities ω_a and σ_a are easily obtainable experimentally and will give a good estimate of the error that can be expected for a given model of operational amplifiers.

The output of the first filter has the transform

$$R_1'(s)=\frac{\omega_a}{s^2+2\sigma_a\omega_a s+\omega_a^2}$$

Therefore, the time response is

$$R_1'(t)=\left(\frac{1}{\sqrt{1-\sigma_a^2}}\right)\times(e^{-\sigma_a\omega_a t}\sin\omega_a\sqrt{1-\sigma_a^2}t)$$

If σ_a is small, then

$$R_1'(t)=(e^{-\sigma_a\omega_a t})\sin\omega_a t$$

The corresponding output of the second filter has the transform

$$C_1'(s)=\frac{\omega_2^2\omega_a}{(s^2+2\sigma_2\omega_2 s+\omega_2^2)(s^2+2\sigma_a\omega_a s+\omega_a^2)}$$

where ω_2 and σ_2 are parameters of the second-order system being tested. The time response is

$$C_1'(t)=\frac{(\omega_2^2e^{-\sigma_a\omega_a t})(\sin(\omega_a\sqrt{1-\sigma_a^2}t-\Omega_a)}{\sqrt{1-\sigma_a^2}(A^2+4AB\sigma_a\omega_a+4B^2\omega_a^2)^{1/2}}+\frac{(\omega_a\omega_2e^{-\sigma_a\omega_a t})(\sin(\omega_2\sqrt{1-\sigma_2^2}t-\Omega_2))}{\sqrt{1-\sigma_2^2}(A^2+4AB\sigma_2\omega_2+4B^2\omega_2^2)^{1/2}}$$

where

$$A=\omega_a^2-\omega_2^2$$

$$B=\sigma_2\omega_2-\sigma_a\omega_a$$

$$\Omega_a=\tan^{-1}\frac{2B\omega_a\sqrt{1-\sigma_a^2}}{-A-2B\sigma_a\omega_a}$$

and

$$\Omega_2=\tan^{-1}\frac{2B\omega_2\sqrt{1-\sigma_2^2}}{-A-2B\sigma_2\omega_2}$$

If $\sigma_2>>\sigma_a$ and σ_a^2 is negligible, then after five time constants of the second-

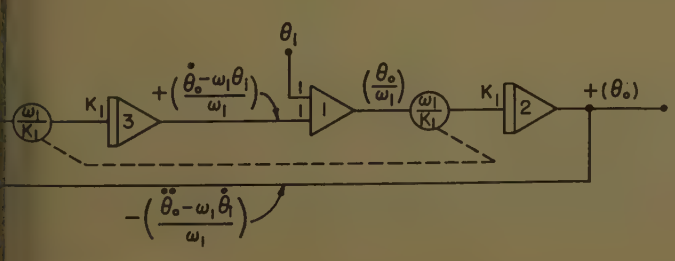


Fig. 7. Computer diagram for synthesis of an ideal filter

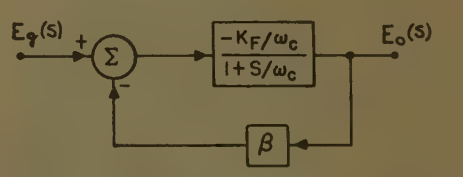
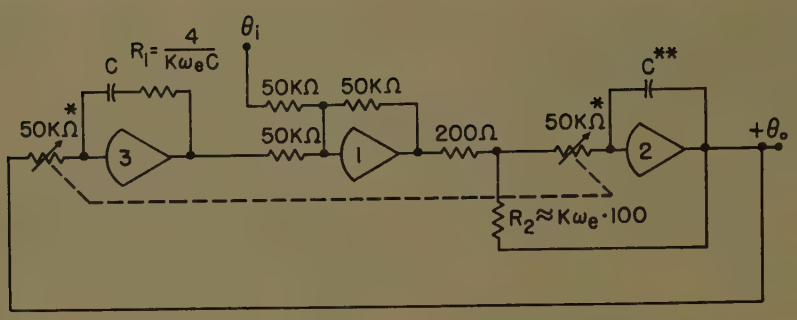


Fig. 8. Essential block diagram of an operational amplifier using positive feedback



* Value at highest measurement frequency

** $C=\frac{1}{(5\cdot10^4)\cdot\omega_{1\,max}}$

Fig. 9. Compensated computer diagram for synthesis of an ideal filter

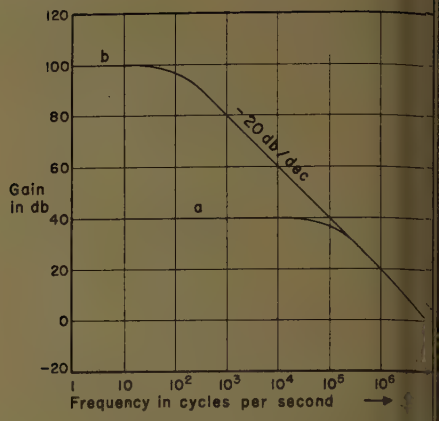


Fig. 10. Graph of response of an operational amplifier: a, with no feedback; b, with controlled positive feedback

order system exponential, the response may be written as

$$C_1'(t)\Big|_{t>\frac{\delta}{\sigma_2\omega_2}} = \frac{\omega_2^2 e^{-\sigma_a\omega_a t}\sin(\omega_a t-\Omega_a)}{[A^2+4AB\sigma_a\omega_a+4B^2\omega_a^2]^{1/2}}$$

where

$$\Omega_a=\tan^{-1}\frac{2B\omega_a}{-A-2B\sigma_a\omega_a}$$

If the ratio of $R_1'(t)$ and $C_1'(t)$ is taken, the magnitude function obtained is

$$\frac{\omega_2^2}{[(\omega_2^2-\omega_a^2)^2+4\sigma_2\omega_a^2\omega_2^2-4\sigma_a\sigma_2\omega_a\omega_2(\omega_a^2+\omega_2^2)]^{1/2}} \qquad (16)$$

This quantity is seen to be the actual transfer function value with an error-causing term in the denominator.

The phase difference in the two functions is

$$\Omega_a=\tan^{-1}\frac{2\sigma_2\omega_a\omega_2-2\sigma_a\omega_a^2}{(\omega_2^2-\omega_a^2)-2\sigma_a\sigma_2\omega_a\omega_2} \qquad (17)$$

This expression has an error-causing term in both the numerator and denominator.

Using the actual magnitude and actual phase difference, a numerical example can be cited to determine the magnitude of the errors. Let it be assumed that the system being studied has the parameters $\sigma_2=0.2$ and $\omega_2=10$ rad/sec. Results of calculations for errors in magnitude are displayed in Table I, the maximum number occurring at resonance. For $\sigma_a=0.005$, the maximum error in magnitude is 2.5%, well within tolerances for most control systems. Similarly, results of calculations for phase errors are displayed in Table II, the maximum occurring above resonance. For $\sigma_a=0.005$, the maximum is 0.5 degree. These figures are given under the assumption that ω_a is the desired measurement frequency and that the two filters are operating at exactly the frequency ω_a. If the two filters are operating at slightly different frequencies, there will be an additional phase error. In practice, this is usually not serious because its total effect can be made less than 1.0 degree. By matching the filters, it can be made negligible.

Thus, with a measuring system damping of 0.005, extreme accuracy is obtainable. Many operational amplifiers are capable of a value of damping of 0.005 to well above 25 rad/sec. Ordinarily, it is not necessary to obtain phase with high accuracy at high frequencies. Under this condition, useful results are obtainable to 250 rad/sec or higher without damping compensation.

Table I. Amplitude Error in Per Cent* $\omega_2=10$ Rad/Sec and $\sigma_2=0.2$

σ_a	ω_a in Rad/Sec 1.00	10.0	100.0	1,000
0.05	0.3	30	0.3	0.03
0.005	0.03	2.5	0.03	0.003
0.0005	0.003	0.3	0.003	0.0003

* A bar under a value indicates that the computed error was slightly less than that shown here.

Equalization

In some control systems it becomes desirable to use a type of compensation when obtaining frequency-domain information by the technique described herein. So doing, accuracy of the data can be assured. In particular, equalization should be used when the transfer function being measured does not have the property given in equation 18:

$$\lim_{s\to 0} G(s)\leq K \qquad (18)$$

where K is finite

Systems not possessing this property have pure integration in the transfer function; that is, at least one s appears as a factor in the denominator. To prevent direct-voltage levels from appearing in the frequency-domain data, differentiate the output of the second filter the same number of times, m, that s can be factored from the denominator of the transfer function.

The low-frequency portion of the response is thus attenuated to easily measurable values. The frequency response obtained from the data with this equalization can then be multiplied by $1/(j\omega_1)^m$ to obtain the actual frequency response of $G(s)$. In other words, one records data on an equalizing curve which increases with frequency and then plays the information back on a complementary curve which decreases with frequency. This technique is very useful when dealing with open-loop transfer functions which contain integration.

The use of one or two differentiations does not present a severe problem because the functions are being measured at low frequencies, allowing the differentiation to be tapered at higher frequencies.

Future Work

An interesting problem in connection with this work concerns logic equipment that must be used with this technique to produce a fully automatic plant observer. This device would have to sense the input function and actuate the observer equipment when the correct input is present.

Table II. Phase Error in Degrees* $\omega_2=1$ Rad/Sec and $\sigma_2=0.2$

σ_a	ω_a in Rad/Sec 1.00	10.0	100.0	1,0
0.05	0.1	3.8	5.7	5.
0.005	0.1	0.3	0.5	0.
0.0005	0.01	0.05	0.05	0.

* A bar under a value indicates that the comput error was slightly less than that shown here.

A second problem, which if solved would tend the usefulness of this technique, that of a statistical study of errors nerated by using inputs which are proximately equal to the test input sired. This problem can probably best handled by assuming that noise is ded to the input functions of the form ven by $R_3(t)$, equation 6. Thus, the ore approximate the input function, e greater is the error. However, the mpling rate is higher, and faster varia- ns in the linear plant parameters can followed.

ummary

A theoretical basis is presented for ob- ining highly accurate complete frequency-domain information of a linear transfer function when it is excited by an impulse, step, or ramp.

This theoretical basis is extended without sacrifice of accuracy to the case where the input is an arbitrary transient of finite duration, and again complete frequency-domain information is obtained. The method is shown to apply equally well to the determination of the Fourier transform of an arbitrary finite-length transient waveform.

This theoretical basis is also shown to be physically realizable, using commercial analog computer components. In particular, the problem of low-frequency analysis of transfer functions is found to be ammenable to this method. An analysis of errors is made when the computer components are not ideal, and the method is found to give high accuracy even under these circumstances.

References

1. Executive-Controlled Adaptive Systems, **R. Staffin, J. G. Truxal.** Polytechnic Institute of Brooklyn, Brooklyn, N. Y., 1958, pp. 1–5, 34–51.

2. Introduction to the Statistical Dynamics of Automatic Control Systems (book), **V. V. Solodovnikov.** Dover Publications, Inc., New York, N. Y., 1960, pp. 35–42.

3. Fourier Integrals for Practical Applications (book), **G. A. Campbell, R. M. Foster.** American Telephone and Telegraph Company, New York, N. Y., 1942, p. 39.

4. Some Limitations on the Accuracy of Electronic Differential Analyzers, **A. B. Macnee.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Mar. 1952, pp. 303–08.

5. Principles of Automatic Controls (book), **F. E. Nixon.** Prentice-Hall, Inc., Englewood Cliffs, N. J., 1953, pp. 371–99.

Determining the Response of Nonlinear Systems to Arbitrary Inputs

RICHARD McFEE
MEMBER AIEE

F AN ENGINEER is presented with a general linear system, a "black box," us say, having an input and an output, e first question he will usually ask is: hat is (or should be) the relationship tween its input and output? This is question which can be answered in many ys. One can determine or prescribe e response of the box to impulses or ep functions or sine waves, and from ese calculate the response to any input, gardless of its character. The concept re is essentially that of impedance, a rm which may be defined in a variety of ys, but the central idea of all the defini- ns is that, in some fashion, the general lationship between the stimulus and e response of a system is specified.

Fixed relationships between input and tput also exist in passive nonlinear sys- ms, at least those whose characteristics e unchanging or which vary systemat- ally with time. Apply a certain input, d some specific response results. If a mputational device is involved which rmits the response to be calculated from arbitrary stimulus, then this device ll have the same conceptual significance impedance in linear systems. And may, with time, develop a practical nificance on a par with the latter.

The theoretical foundation for a concept nonlinear impedance now exists. Wiener,[1] Singleton,[2] Zadeh,[3] Bose,[4] Boonton,[5] Blackman,[6] and others have presented a variety of general theorems regarding nonlinear systems which may be considered in this light. They show that most of these systems can be represented within arbitrarily small error by certain infinite series.

Of these approaches the viewpoint of Singleton and Zadeh is perhaps most compatible with the background and philosophy of the engineer. Singleton shows that the response of a passive, invariant, nonlinear system to an arbitrary input can be computed in terms of its response to impulses. The procedure is analogous to the use of impulse response in linear systems. Zadeh expresses Singleton's result in a concise mathematical form, by means of a series of integrals of the convolution type, the first term of which is the familiar superposition integral of linear systems. Their approach actually originated with Volterra. It is particularly appropriate to slightly nonlinear systems such as parametric amplifiers and it can be applied to circuits such as flip-flops, and more exotic devices such as subharmonic generators.

In this article the approach of Singleton and Zadeh is extended in two ways. First an explicit method is developed for determining the kernels of the integral of Zadeh's series. Second, it is shown that these kernels can be represented by equivalent circuits containing linear networks and multipliers. Blackman[6] has proved that such equivalents exist in general. These equivalent circuits provide the engineer with an insight into the functioning of the nonlinear system which is not easily obtained from the equations. To illustrate the principles involved, two simple examples are given.

The Response Functions

Fig. 1 shows a sketch of a typical input to a nonlinear system, and an approximation to it in terms of a large number of short rectangular pulses of a uniform width Δt, one adjacent to the next. The time t shown on the graph represents the moment at which the output is to be determined. It is a basic assumption here that replacement of the actual input with sufficiently fine-grained approximation of it, such as is typified by Fig. 1(B), will change the output by a negligible amount. This assumption is met by all real systems, where the response at some high frequency eventually falls to zero, but it is not met by some of the ideal circuits, such as a perfect differentiator. It is also assumed that the circuit is stable and passive, that it does not change with time, and that it has a finite memory. The input to the system can be described in terms of the heights of the various pulses shown in Fig. 1(B). If these are

Paper 61-114, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 29–February 3, 1961. Manuscript submitted June 8, 1960; made available for printing November 23, 1960.

Richard McFee is with Syracuse University, Syracuse, N. Y.

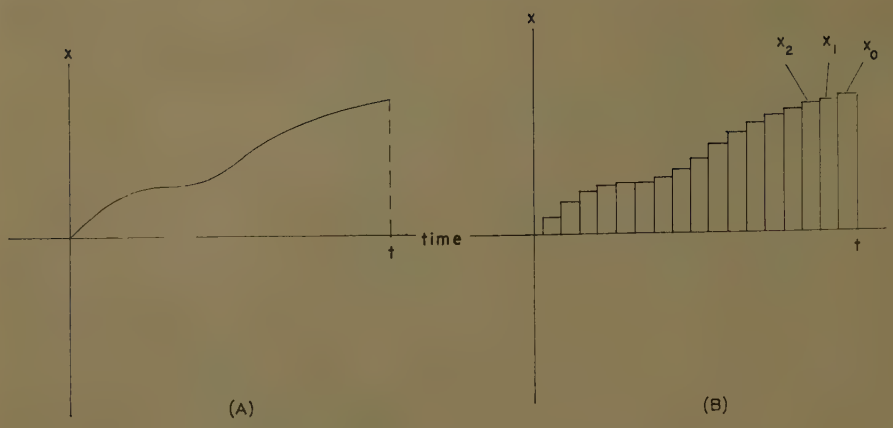


Fig. 1. Multiple-pulse approximation of input functions

A—Input B—Multipulse approximation

labelled x_0, x_1, etc., as shown in Fig. 1, then the output y may be considered a function of many variables, i.e.,

$$y=f(x_0, x_1, \ldots x_n) \tag{1}$$

A function of many variables can be represented by a multidimensional Maclaurin series. Although the conditions for such a representation can be made quite weak (the Wierstrass-Stone theorem), it will be assumed here that all derivatives are continuous and that the remainder of the series goes to zero as the number of terms becomes infinite. This simplifies considerably the problem of finding the kernels of the integrals of Zadeh's series. The more complex general problem will not be considered here.

The multidimensional Maclaurin series has the form

$$\begin{aligned} y&=(a_0x_0+a_1x_1+\ldots a_nx_n)+\\ &\quad (a_{00}x_0^2+a_{01}x_0x_1+\ldots a_{nn}x_n^2)+\ldots\\ &=\sum_{k=0}^{n} a_kx_k+\sum_{k=0}^{n}\sum_{j=0}^{n} a_{kj}x_kx_j+\ldots\\ &=y_L+y_Q+y_C+\ldots \end{aligned} \tag{2}$$

where y_L is the linear term, y_Q the quadratic term, etc. The number n represents the number of variables, and is determined by the granularity of the multipulse approximation to the input as well as by the length of the "memory" of the system.

If the coefficients of all the terms of second degree or higher are zero, then the series becomes

$$y=y_L=a_0x_0+a_1x_1+\ldots+a_nx_N \tag{3}$$

which is a discrete approximation to the superposition integral for linear circuits. This may be seen by defining new coefficients $h_0, h_1, \ldots h_N$ by

$$h_k=a_k/\Delta \tag{4}$$

where Δ is the length of each of the pulses. Substituting in equation 3 gives

$$y_L=h_0x_0\Delta+h_1x_1\Delta+\ldots h_nx_n\Delta \tag{5}$$

If Δ is chosen sufficiently small the above sum approaches

$$y_L=\int_0^t h_L(\lambda)x(t-\lambda)d\lambda \tag{6}$$

where t is $n\Delta$. This is the familiar superposition integral of linear systems. It is assumed implicitly in this equation that the excitation starts when t is zero.

The coefficients of equation 3 represent the output of the system for a pulse applied at various times in the past. For example, the term a_0 represents the response to a unit pulse of width Δ which has just terminated, while the term a_2 represents the response to a pulse which has terminated 2Δ seconds ago. The sequence a_0, a_1, a_2, etc., represents the system response to a single applied pulse which is "receding into the past" and the sequence is a discrete approximation to the impulse response which is $h_L(\lambda)$.

In nonlinear systems, the higher-order terms of equation 2 are not zero. The coefficients of the higher-order terms of a nonlinear circuit can be determined in the same way as the coefficients of the equation for a linear system. A very weak pulse is applied at first and from the response to it, the linear part of the network response due to the coefficients a_0, a_1, a_2, etc., is determined. The strength of the applied pulse is then increased, and the higher-order coefficients such as a_{∞} and a_{33} determined.

To find the coefficients of the cross-product terms, such as a_{15}, it is necessary to apply two pulses at different times and to determine the deviation of the response from the response which results from applying the pulses one at a time. Mathematically speaking, the coefficient a_{15} is $\partial^2y/\partial x_1\partial x_5$, and similar formulas apply to all other coefficients. In this way all the coefficients can be determined from the response of the system to single and asynchronous groups of pulses. Then, when all the coefficients have been found, the response of the network to any input, as approximated by the series of rectangular pulses x_0, x_1, x_2, etc., can be calculated by substituting into equation 2.

To express Y_Q in equation 2 in integral form we define the coefficients h_{kj} by

$$h_{kj}=\frac{a_{kj}}{\Delta^2}$$

and substitute into the quadratic term of equation 2, with the result:

$$y_Q=[h_{00}x_0^2+h_{01}x_0x_1+\ldots h_{nn}x_n^2]\Delta^2$$

If the Δ have been made small enough so that neither h_{kj} or x_k change very much from one term to the adjacent one, the series is represented by the integral

$$y_Q=\int_0^t\int_0^t h_Q(\lambda_1, \lambda_2)x(t-\lambda_1)x(t-\lambda_2)\times d\lambda_1d\lambda_2$$

The cubic terms y_C of the expansion are represented by the integral

$$y_C=\int_0^t\int_0^t\int_0^t h_C(\lambda_1,\lambda_2,\lambda_3)x(t-\lambda_1)\times x(t-\lambda_2)x(t-\lambda_3)d\lambda_1d\lambda_2d\lambda_3$$

Continuing to quartic and even higher order terms leads to the general equation for the output of a nonlinear system

$$y=y_L+y_Q+y_C+y_{QR}+\ldots$$

where y_{QR} are defined by equations similar to 6, 9, and 10. The sequence functions $h_L(\lambda)$, $h_Q(\lambda_1, \lambda_2)$, $h_C(\lambda_1, \lambda_2, \lambda_3)$ may be designated as the response functions of the system.

The response functions in the representation 11 are defined by equation itself. Let the input be a very short pulse of unit amplitude and width occurring at a time λ_0 seconds before the instant t of measurement. Substitution into equation 11 gives as the output

$$y(t)=h_L(\lambda_0)\Delta+h_Q(\lambda_0, \lambda_0)\Delta^2+\ldots$$

The function $h_L(\lambda)$ is the limit of the ratio of the output to Δ as it approaches zero, i.e., it is $\partial y/\partial\Delta$. In the same way $h_Q(\lambda_0, \lambda_2)$ is $1/2\ (\partial^2y/\partial\Delta^2)$. If two asynchronous pulses are applied at λ_1 and λ_2 sec (seconds) before the time t, both with unit amplitude and width Δ_1 and the subsequent response will be

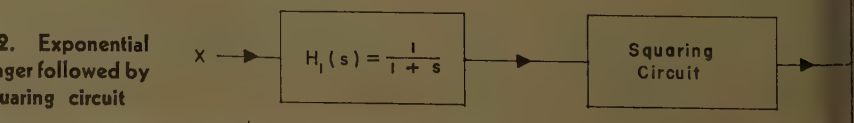


Fig. 2. Exponential averager followed by squaring circuit

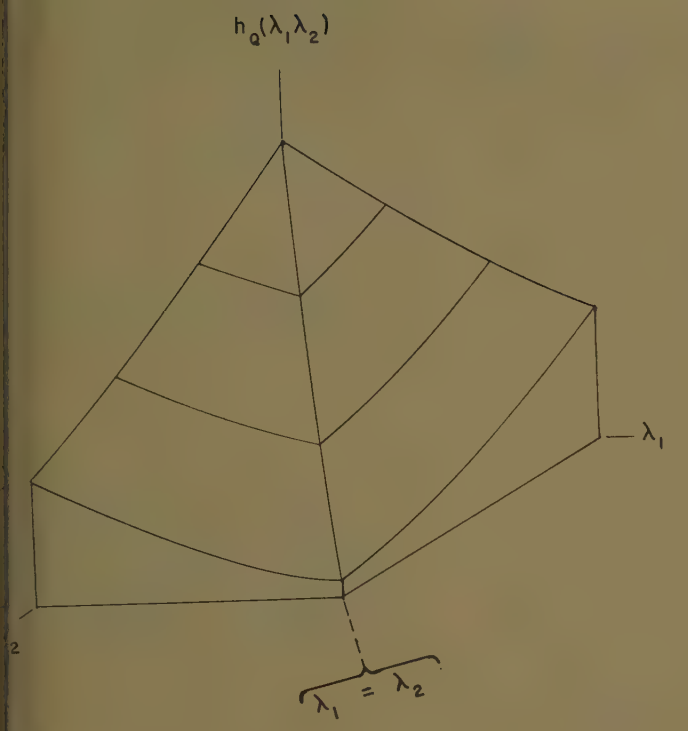


Fig. 3. Surface representing quadratic response function $e^{-\lambda_1-\lambda_2}$. Quadratic component y_Q system output y given by integral of product of this surface and input function $x(t-\lambda_1)$ $x(t-\lambda_2)$ over region $\lambda_1<t$, $\lambda^2<t$ as specified by equation 9

$t) = h_L(\lambda_1)\Delta_1 + h_L(\lambda_2)\Delta_2 + h_Q(\lambda_1, \lambda_1)\Delta_1^2 + 2h_Q(\lambda_1 \lambda_2)\Delta_1\Delta_2 + h_Q(\lambda_2, \lambda_2)\Delta_2^2$ (13)

he cross-product term $h_Q(\lambda_1, \lambda_2)$ is 1/2 $^2y/\partial\Delta_1\,\partial\Delta_2$).

Subtracting out the responses that ould result if the pulses were impressed one will leave only this cross-product rm multiplied by $2\Delta_1\Delta_2$. By applying ulses in this manner, either physically or nalytically, the response functions can determined. Specific examples of the ocedure will be given presently.

The higher-order terms of the response nctions have properties analogous to e properties of the impulse response. the output of a linear system contains component directly proportional to the put, then the impulse response $h_L(\lambda)$ ll itself contain an impulse. In the me way, if the output of a nonlinear stem contains a component directly oportional to the square of the input, e quadratic term of the response nction $h_Q(\lambda_1, \lambda_2)$ will contain a 2-dimenonal impulse, i.e.,

$\lambda_1, \lambda_2) = p\delta(\lambda_1, \lambda_2)$ (14)

here p is a constant and

$\lambda_1, \lambda_2) = 0$ for $|\lambda_1|>0$, $|\lambda_2|>0$ (15)

d

$\int_0^t \delta(\lambda_1, \lambda_2)d\lambda_1 d\lambda_2 = 1$

r example, a simple squaring circuit s the response function

$(\lambda_1 \lambda_2) = \delta(\lambda_1 \lambda_2)$ (16)

l other terms are zero.

Systems With Several Inputs

Equation 11, just obtained for a system with one input, can be generalized to take into account systems with several inputs. Parametric amplifiers fall into this class. It is not necessary that the inputs have the same character: one input can be a voltage, the other a shaft rotation, the third a hydraulic pressure. We will consider here only systems with two inputs; the extension to a greater number should be clear from the approach.

If a nonlinear system has two inputs, each may be represented by an approximation of the type shown in Fig. 1. If one of the inputs is designated as $x_0, \ldots x_n$ and the other as $z_0, \ldots z_n$, then the output can be expressed as

$$f(x_0 \ldots x_n, z_0 \ldots z_n) \quad (17)$$

Expansion of this into a multidimensional Maclaurin series, and expression of this series in integral form, gives

$$\begin{aligned} f(t) = {} & \int_0^t h_{Lx}(\lambda)x(t-\lambda)d\lambda + \\ & \int_0^t h_{Lz}(\lambda)z(t-\lambda)d\lambda + \\ & \int_0^t \int_0^{\lambda_2} h_{Qx}(\lambda_1, \lambda_2)x(t-\lambda_1)x(t- \\ & \lambda_2)d\lambda_1 d\lambda_2 + \int_0^t \int_0^{\lambda_2} h_{Qxz}(\lambda, \lambda_2)\times \\ & x(t-\lambda)z(t-\lambda_1)d\lambda_1 d\lambda_2 + \\ & \int_0^t \int_0^{\lambda_2} h_{Qz}(\lambda_1, \lambda_2)z(t-\lambda_1)\times \\ & z(t-\lambda_2)d\lambda_1 d\lambda_2 + \ldots \end{aligned} \quad (18)$$

The output is a superposition of responses to each input by itself, plus components representing the cross product or "intermodulation" of the two inputs.

A simple example of a nonlinear system with two inputs is a multiplier. Its response function is $h_{Qxz}(\lambda_1, \lambda_2) = \delta(\lambda_1, \lambda_2)$.

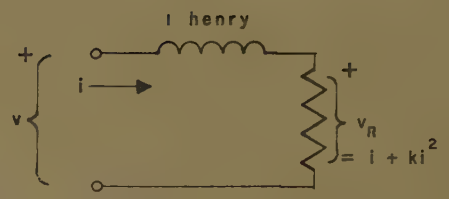


Fig. 4. Inductor with nonlinear resistor

Equivalent Circuits

If two systems have the same response functions h_L, h_Q, h_C ..., they are equivalent. The same input will produce the same output from both. The equivalence can be partial. If both have the same h_L functions, they will be equivalent in the linear component of their response. Small inputs will be handled in a similar way by both. If h_Q of both are also the same, the linear and the quadratic response term will be identical. The equivalence of the response will hold for still larger signals. The equivalence becomes more exact as more and more terms are included. For many engineering purposes, it is sufficient to include only the linear and quadratic terms.

Example 1. Exponential Averager With Squaring Circuit

The problem is to determine the step response of a circuit consisting of an *R-C* (resistance-capacitance) network with impulse response e^{-t} followed by a squaring circuit; see Fig. 2. The unit step is applied at time equal to zero.

To find the response function we apply two short pulses of unit amplitude, first, one of length Δ_f, then, a second of length Δ_s at λ_f and λ_s sec respectively before the time $t$when the output is measured. The output y_p produced by these pulses is

$$y_p = (\Delta_f e^{-\lambda_f} + \Delta_s e^{-\lambda_s})^2 \quad (19)$$

or

$$y_p = (\Delta_f)^2 e^{-2\lambda_f} + 2\Delta_f\Delta_s e^{-\lambda_f-\lambda_s} + (\Delta_s)^2 e^{-2\lambda_s} \quad (20)$$

There is no linear term and hence $h_L(\lambda)$ is zero. If three or more pulses had been applied, all terms would still be of the type $(\Delta_n)^2$ or $\Delta_n\Delta_m$, and therefore $h_Q(\lambda_1, \lambda_2, \lambda_3)$ and higher-order terms are also zero. Computing $1/2(\partial^2 y/\partial\Delta_f^2)$ and $1/2$ $(\partial^2 y 2/\Delta_f\partial\Delta)_s$, reveals that

$$h_Q(\lambda_1, \lambda_2) = e^{-\lambda_1-\lambda_2} \quad (21)$$

The output of the circuit with a step applied is therefore

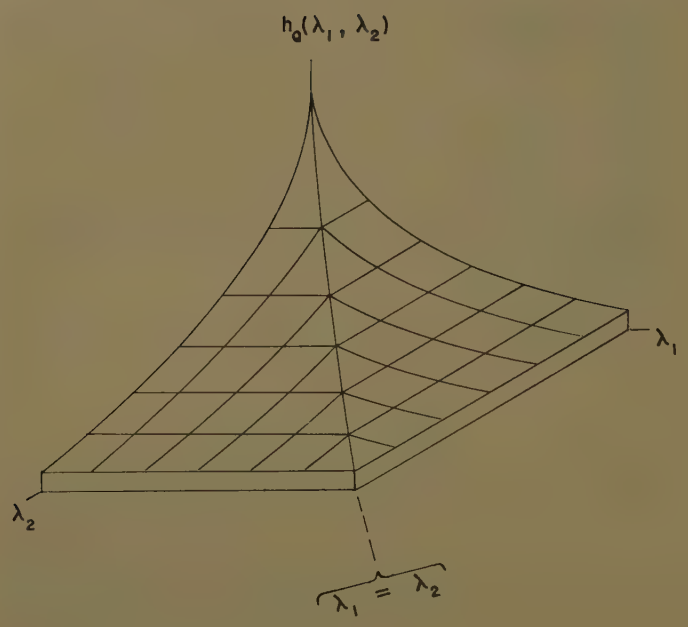


Fig. 5. Surface representing quadratic response functions $e^{-\lambda_2}$ for $\lambda_2 > \lambda_1$ and $e^{-\lambda}$ for $\lambda_1 > \lambda_2$

$$y = \int_0^t \int_0^t e^{-\lambda_1-\lambda_2} u(t-\lambda_1)u(t-\lambda_2)d\lambda_1 d\lambda_2$$
$$= 2\int_0^t \int_0^{\lambda_2} e^{-\lambda_1-\lambda_2} u(t-\lambda_1)u(t-\lambda_2)d\lambda_1 d\lambda_2$$
$$= 2\int_0^t \int_0^{\lambda_2} e^{-\lambda_1-\lambda_2} d\lambda_1 d\lambda_2 \qquad (22)$$

where $u(t)$ is the unit step function, which is 1 for t greater than zero, and zero for t less than zero. Evaluation of the above integral gives

$$y = e^{-2t} - 2e^{-t} + 1 = (e^{-t}-1)^2 \qquad (23)$$

This answer could, of course, have been worked out in a much simpler way. The step response of a circuit with an impulse response of e^{-t} is

$$v = (1-e^{-t}) \qquad (24)$$

The output of the circuit is the square of this, i.e., the same result given in equation 23. The convolution series is used only to introduce the essential ideas of the analysis with a simple example. The $h_Q(\lambda_1, \lambda_2)$ function may be represented geometrically as is shown in Fig. 3 for equation 21.

Example 2. Inductor With Nonlinear Resistor

A unit step of voltage is applied to a circuit consisting of a linear 1-henry inductor in series with a slightly nonlinear resistor; see Fig. 4. The current i to be determined. The resistor is characterized by

$$v_r = i + ki^2 \qquad (25)$$

The resistance acts a little like a rectifier. It presents higher impedance to current flowing in one direction than the other. The nonlinear portion of the resistance expressed by the factor k is assumed to be small.

The conventional solution will be obtained first. The differential equation of the circuit, after a 1-volt step has been applied is

$$1 = \frac{di}{dt} + i + ki^2 \qquad (26)$$

When this equation is integrated, the value of the current is found to be

$$i = \left(\frac{\sqrt{1+4k}-1}{2k}\right) \frac{1-e^{-t\sqrt{1+4k}}}{1+\dfrac{\sqrt{1+4k}-1}{\sqrt{1+4k}+1}e^{-t\sqrt{1+4k}}} \qquad (27)$$

When k is small, we can make the substitutions

$$(\sqrt{1+4k}-1)/2k \simeq 1-k$$
$$\sqrt{1+4k} \simeq 1+2k$$
$$(\sqrt{1+4k}-1)/\sqrt{1+4k}+1) \simeq k$$
$$e^{-t(1+2k)} \simeq e^{-t}(1-2kt) \qquad (28)$$

and equation 27 becomes

$$i = (1-e^{-t}) - k(1-2te^{-t}-e^{-2t}) \qquad (29)$$

where the first two terms represent the solution which would be obtained if the resistance were linear, i.e., if the factor k were zero. The last three terms represent the perturbing influence of the nonlinearity on the output.

Now, we will solve for the output of the circuit using response functions. Since we are considering the situation when the nonlinearity of the circuit is small, it is necessary to evaluate only $h_L(\lambda)$ and $h_Q(\lambda_1, \lambda_2)$. To do this we consider the response of the circuit first to one pulse alone, and then its response to a group of two pulses.

If a pulse of unit voltage and length Δ is applied, its effect is to produce an initial current i equal to the length of t pulse divided by the inductance. Sini the latter is one henry, the initial curre is simply the length of the pulse Δ. T differential equation of the circuit aft this pulse has been applied is

$$0 = \frac{di}{dt} + i + ki^2 \qquad (3$$

where the current i_0 is equal to Δ at t equ to zero. The solution to this equation

$$i = \frac{1}{\left(\frac{1}{i_0}+k\right)e^t - k} = \frac{1}{\left(\frac{1}{\Delta}+k\right)e^t - k} \qquad (3$$

We now apply two pulses of unit amp tude, the first of length Δ_f and occurri λ_f sec before the time t, and the seco of length Δ_s occurring λ_s sec before t time t. Just prior to the time the seco pulse is applied, the current $i(t-\lambda_s+)$ the inductor due to the first pulse is

$$i(t-\lambda_s+) = \frac{1}{\left(\frac{1}{\Delta_f}+k\right)e^{\lambda_f-\lambda_s} - k} \qquad (3$$

The current $i(t-\lambda_s-)$ immediately aft the second pulse is this current plus t added currents Δ_s, giving a total

$$i(t-\lambda-) = \Delta_s + \frac{1}{\left(\frac{1}{\Delta_f}+k\right)e^{\lambda_f-\lambda} - k} \qquad (3$$

which decays according to equation Consequently the current at time t will

$$i(t) = \frac{1}{\left[\dfrac{1}{\Delta_s + \left(\dfrac{1}{\left[\frac{1}{\Delta_f}+k\right]e^{\lambda_f-\lambda_s}-k}\right)} + k\right]e^{\lambda_s}}$$

The equation is for $\lambda_f > \lambda_s$. Expandi it in a power series in k, keeping only fir order terms, yields

$$i(t) = (\Delta_f e^{-\lambda_f} + \Delta_s e^{-\lambda_s}) + k(\Delta_f^2 e^{-2\lambda_f} + 2\Delta_f\Delta_s e^{-\lambda_f-\lambda_s} + \Delta_s^2 e^{-2\lambda_s}) - k(\Delta_f^2 e^{-\lambda_f} + 2\Delta_f\Delta_s e^{-\lambda_f} + \Delta_s^2 e^{-\lambda_s}) \qquad (3$$

Computing $\partial i/\partial\Delta_f$, $1/2\ (\partial^2 i/\partial\Delta_f\partial\Delta$ etc., leads directly to the response fun tions

$$h_L(\lambda) = e^{-\lambda}$$
$$h_Q(\lambda_1, \lambda_2) = h_{Qa} + h_{Qb} \qquad (3$$

where

$$h_{Qa} = ke^{-\lambda_1-\lambda_2}$$
$$h_{Qb} = -ke^{-\lambda_2} \text{ for } \lambda_2 > \lambda_1$$
$$= -ke^{-\lambda_1} \text{ for } \lambda_1 > \lambda_2$$

The h_{Qa} function is shown pictorially Fig. 3 and the h_{Qb} function in Fig. 5.

The output of the circuit for a u

p applied at t equal to zero is therefore
en by the integrals

$$=\int_0^t (e^{-\lambda})u(t-\lambda)d\lambda+$$
$$k\int_0^t\int_0^t e^{-\lambda_1-\lambda_2}u(t-\lambda_1)u(t-\lambda_2)d\lambda_1d\lambda_2+2k\int_0^t\int_0^{\lambda_2} e^{-\lambda_2}u(t-\lambda_1)u(t-\lambda_2)d\lambda_1d\lambda_2 \quad (37)$$

We have made use here of the symtry of the h_{Qb} function about the line $=\lambda_2$. The last integration is over half the positive quadrant of the λ_1, λ_2 ne. The result after integration is

$$\simeq i_1+i_Q=(1-e^{-t})-k(1-2te^{-t}-e^{-2t}) \quad (38)$$

ch is the same result as found in equa- 29 by the conventional method. lowever, in addition to determining output of the circuit when a unit step pplied, we have also determined the onse functions which apply to *any* it. Thus if the input is $v(t)$, the out- will be

$$\simeq\int_0^t e^{-\lambda}v(t-\lambda)d\lambda+k\int_0^t\int_0^t e^{-\lambda_1-\lambda_2}v(t-\lambda_1)v(t-\lambda_2)d\lambda_2d\lambda_1+2k\int_0^t\int^{\lambda_2} e^{-\lambda_2}v(t-\lambda_1)v(t-\lambda_2)d\lambda_1d\lambda_2 \quad (39)$$

s is a much more general result than obtained by the conventional method. his equation can be represented by equivalent circuit. The first term i_2 hat of an *R-C* integrator, which is esented by the transfer function $+s$. The second term i_{Qa} is the same nat of the *R-C* integrator followed by ultiplier which is discussed in example

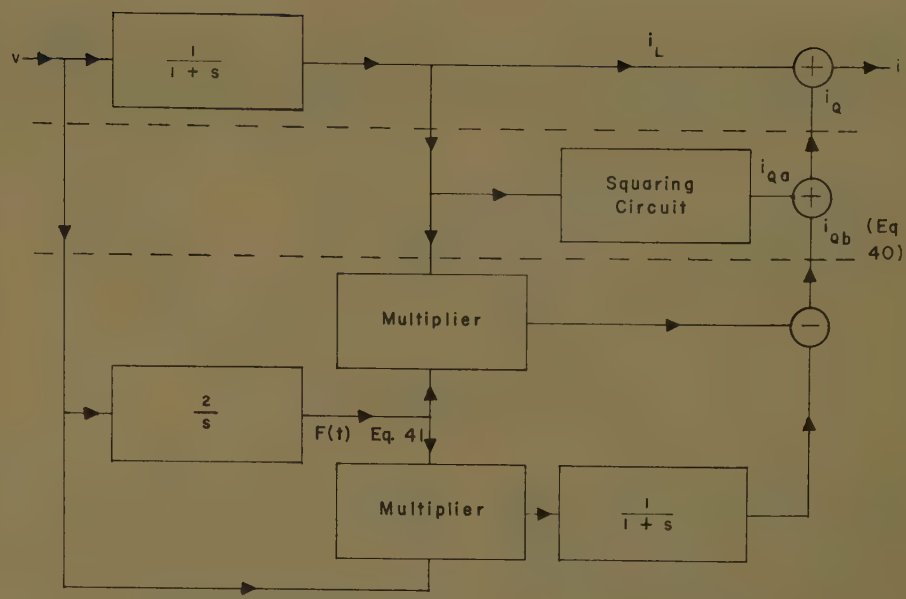


Fig. 6. Equivalent circuit of nonlinear circuit shown in Fig. 4

1. The third term may be rewritten as

$$i_{Qb}=\int_0^t e^{-\lambda}[F(t)-F(t-\lambda)]v(t-\lambda)d\lambda = F(t)\int_0^t e^{-\lambda}v(t-\lambda)d\lambda-\int_0^t e^{-\lambda}F(t-\lambda)v(t-\lambda)d\lambda \quad (40)$$

where $F(t)$ is defined by the equation

$$F(t)=2\int_0^t v(t-\lambda)d\lambda \quad (41)$$

Equation 40 is represented by the equivalent circuit of Fig. 6, where the overall equivalent circuit of the nonlinear network is also shown.

References

1. Nonlinear Problems in Random Theory (book), **N. Wiener.** John Wiley & Sons. Inc., New York, N. Y., 1958.

2. Theory of Nonlinear Transducers, **H. E. Singleton.** *Technical Report No. 160* Massachusetts Institute of Technology, Cambridge, Mass., 1950.

3. Nonlinear Multipoles, **L. A. Zadeh.** *Proceedings*, National Academy of Sciences, Wash ington, D. C., vol. 39, 1953, pp. 274–80.

4. A Theory of Nonlinear Systems, **A. Bose.** *Report No. 309*, Massachusetts Institute Technology, 1956.

5. The Measurement and Representation of Nonlinear Systems, **R. C. Boonton.** *Transactions*, Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., CT-1, 1954, pp. 32–34.

6. The Representation of Non-linear Networks, **J. Blackman.** *Report No. 81560*, Syracuse University Research Institute, Syracuse, N. Y. (To be published.)

7. Theorie Generale des Fonctionelles (book), **V. Volterra, J. Peres.** Libraire du Bureau des Longetudes, Ecole Polytechnique, Paris, France, vol. I, 1936, chap. IV.

On Linear Control Theory

PETER D. JOSEPH
STUDENT MEMBER AIEE

JULIUS T. TOU
MEMBER AIEE

psis: Optimum control of linear ivariable processes is the concern of paper. The optimum control problem udied via the state-variable concept the dynamic-programing technique. near multivariable control system is n to have optimum performance with ct to a quadratic performance cri- n, if both the controller and the ator are independently designed in an um manner.

R ABOUT a quarter of a century, he conventional linear control theory njoyed a remarkable rate of growth. r years of rapid and active development, it appears that the conventional approach of control-system design is reaching its full capability. As a result, control scientists and engineers in this country and in the Soviet Union have launched the exploration of a new philosophy, the search for novel techniques and the development of new theory for control-system design. The research in this direction has been accelerated by the aid of modern high-speed digital computers, and inspired by the inauguration of new branches of applied mathematics, such as dynamic programing. In recent years, optimum control has become a subject of much interest, a number of papers concerning it having appeared in the literature.

An optimum multivariable control system is generally equipped with an optimum controller and an estimator. The controller is used to generate an optimum control sequence; and the estimator is employed to determine all the state variables from the measurable output variables of the process. The functions of both the controller and the

Paper 61-728, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Ithaca, N. Y., June 18–23, 1961. Manuscript submitted January 30, 1961; made available for printing April 10, 1961.

Peter D. Joseph and Julius T. Tou are with Purdue University, Lafayette, Ind.

The work reported here was supported by the Office of Naval Research under Contract *Nonr-1100(18)*, through the Information Systems Branch.

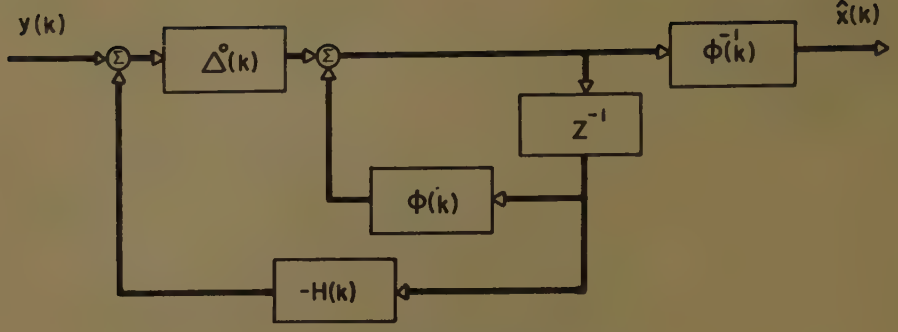


Fig. 1. Optimum estimator when $\mathbf{m}=0$

estimator may be performed by a modern high-speed digital computer. In 1958, Kalman and Koepcke[1] raised the question of "whether the separate optimization of statistical prediction and control-system performance yields a system which is optimal in the *over-all* sense." This paper answers the question in the affirmative for an important class of control problems.

A linear discrete time plant[2] obeys a vector difference equation

$$\mathbf{x}(k+1)=\Phi(k)\mathbf{x}(k)+\mathbf{A}(k)\mathbf{m}(k) \qquad (1)$$

where $\mathbf{x}(k)$ is the state vector, $\mathbf{m}(k)$ is the vector whose components are the control signals, $\Phi(k)$ is the state-transition matrix, and $\mathbf{A}(k)$ determines how the control signal affects the state of the plant.[1-5]

In many problems, it is desirable to choose $\mathbf{m}(k)$ so as to minimize a quadratic performance index of the form

$$I_K=\sum_{k=1}^{K}\mathbf{x}'(k)\mathbf{Q}\mathbf{x}(k) \qquad (2)$$

for any initial $\mathbf{x}(0)$. In equation 2, $\mathbf{Q}$ is a positive definite matrix and the prime denotes the transpose of a vector, or a matrix. The use of dynamic programing[6] allows the determination of the optimum control law.[1,7] This law has two important properties.

a. The optimum control signal $\mathbf{m}^\circ(k)$ is a linear function of $\mathbf{x}(k)$, i.e.

$$\mathbf{m}^\circ(k)=\mathbf{B}^\circ(k)\mathbf{x}(k) \qquad (3)$$

where $\mathbf{B}^\circ(k)$ is a matrix which can be determined.

b. If the plant is subjected to additive random *white* noise disturbances and the optimum control law, equation 3, is used, the system is still optimal in the sense that EI_K is minimized, where E is the expected value operator.[7]

In order to use equation 3 to generate the control signals, all components of $\mathbf{x}(k)$ must be measurable. In many practical cases, however, not all of the state variables are directly measurable. In more quantitative terms, only certain signals are directly measurable. These signals will be called the output signals,[2,3] and a linearly independent set of them will be denoted by $y_1(k)$, $y_2(k)\ldots$. For a linear plant, the output signals are linear functions of the state variables, $\mathbf{x}(k)$, i.e.

$$\mathbf{y}(k)=\mathbf{H}(k)\mathbf{x}(k) \qquad (4)$$

where $\mathbf{y}(k)$ is the vector whose components are $y_1(k)$, $y_2(k)\ldots$.[2,3] The interesting case occurs when there are fewer output signals than state variables. Then $\mathbf{H}(k)$ is not a square matrix and cannot be inverted.

In order for $\mathbf{m}(k)$ to be *realizable*, it must be a function only of the measurable outputs $\mathbf{y}(k)$, $\mathbf{y}(k-1)\ldots$. But the optimum $\mathbf{m}^\circ(k)$ is a function of $\mathbf{x}(k)$. The essential problem then is what is the optimum *realizable* $\mathbf{m}(k)$.

Optimum Estimation

The problem of estimating $\mathbf{x}(k)$ from a knowledge of $\mathbf{y}(k)$, $\mathbf{y}(k-1)\ldots$ has been considered in various forms by many authors. The particular formulation of interest here is due to Kalman.[3,8] The optimum estimate of $\mathbf{x}(k)$ is defined to be the one that minimizes the expected value of the square of the Euclidian norm of $\mathbf{x}(k)-\hat{\mathbf{x}}(k)$, where the estimate is denoted by $\hat{\mathbf{x}}(k)$. When $\mathbf{m}(k)=\mathbf{m}(k-1)=\ldots=0$, the optimum $\hat{\mathbf{x}}(k)$ can be generated by the system of Fig. 1 for a plant subject to a random *white* noise disturbance.[8] The system consists of a copy of the plant dynamics, to which a correction signal is added. This correction signal is generated by operating on the difference between $\mathbf{H}(k)\Phi(k-1)\hat{\mathbf{x}}(k-1)$ (the "best" estimate of $\mathbf{y}(k)$ given $\mathbf{y}(k-1)$, $\mathbf{y}(k-2)$, ...) and the true value of $\mathbf{y}(k)$ with a matrix $\Delta^\circ(k)$. The optimum $\Delta^\circ(k)$ can be determined by the methods of reference 8.

The optimum estimator has four important properties:[8]

1. Let $\mathbf{x}(k)-\hat{\mathbf{x}}(k)=\tilde{\mathbf{x}}(k)$. Let $Y(k)$ be the set of all random vectors that can be expressed as a linear function of $\mathbf{y}(k)$, $\mathbf{y}(k-1)\ldots$. Then $\tilde{\mathbf{x}}(k)$ is orthogonal to every vector in $Y(k)$, i.e., if $\mathbf{a}$ is in Y then $E[\mathbf{a}'\tilde{\mathbf{x}}(k)]=0$.

2. The estimator of Fig. 1 is the optimum linear estimator.

3. If the disturbance is gaussian, it is the optimum estimator.

4. If $\mathbf{m}^\circ(k)$ is given by equation 3, the realizable estimate of $\mathbf{m}^\circ(k)$, $\hat{\mathbf{m}}^\circ(k)$, which minimizes the expected value of the square of the Euclidian norm of $\mathbf{m}^\circ(k)-\hat{\mathbf{m}}^\circ(k)$ is given by

$$\hat{\mathbf{m}}^\circ(k)=\mathbf{B}^\circ(k)\hat{\mathbf{x}}(k)$$

If $\mathbf{m}(k)$, $\mathbf{m}(k-1)\ldots$ are not zero, it is easy to show that the optimum estimate is determined by the system of Fig. 2.

When $\mathbf{m}(k)$, $\mathbf{m}(k-1)\ldots=0$, Fig. 1 is the same as Fig. 2 and therefore $\mathbf{x}(k)-\hat{\mathbf{x}}(k)$ is the same $\hat{\mathbf{x}}(k)$ discussed earlier. When the disturbances and the initial value of $\mathbf{x}(0)$ equals zero, but $\mathbf{m}(k)$, $\mathbf{m}(k-1)\ldots$ does not equal zero, it is seen by inspection that $\mathbf{x}(k)=\hat{\mathbf{x}}(k)$. Thus, by superposition $\mathbf{x}(k)-\hat{\mathbf{x}}(k)=\tilde{\mathbf{x}}(k)$ and Fig. 2 must be the optimum estimator. Note that $\tilde{\mathbf{x}}(k)$ is independent of $\mathbf{m}(j)$ for all j (Property 5).

The Over-All System

The question now is whether the optimum control law, equation 3, when the optimum estimator, will yield an optimum over-all system. The answer is yes.

To state this more precisely, we are given a linear discrete time plant, let $\mathbf{B}^\circ(k)$ be a sequence of matrices so that the control signal

$$\mathbf{m}^\circ(k)=\mathbf{B}^\circ(k)\mathbf{x}(k)$$

minimizes

$$I_K=\sum_{k=1}^{K}\mathbf{x}'(k)\mathbf{Q}\mathbf{x}(k)$$

for any arbitrary initial condition.

Given the same linear discrete time plant, whose state variables now are not all measurable and which is subject to a white gaussian disturbance, let $\hat{\mathbf{x}}(k)$ be the vector in $Y(k)$, the set of all linear combinations of the past outputs $y(k)$, $y(k-1)\ldots$ which minimizes $E[||\mathbf{x}(k)-\hat{\mathbf{x}}(k)||^2]$, where $||\ ||$ denotes the Euclidian norm.

Theorem

Given a linear discrete time plant whose state variables are not all measurable and which is subject to an additive white gaussian disturbance, the realizable control signal which minimizes

$$_K = E\sum_{k=1}^{K} \mathbf{x}'(k)\mathbf{Q}\mathbf{x}(k) \tag{8}$$

given by

$$^\circ(k) = \mathbf{B}^\circ(k)\hat{\mathbf{x}}(k) \tag{9}$$

Proof: Assume momentarily that the turbance is zero and let

$$= \sum_{j=K-N}^{K} \mathbf{x}'(k)\mathbf{Q}\mathbf{x}(k) \tag{10}$$

so let

$$= J_N \tag{11}$$

en $\mathbf{m}(k) = \mathbf{m}^\circ(k)$

the principle of optimality, f_N is absolute minimum of J_N. If there no disturbance, $J_0 = \mathbf{x}'(K)\mathbf{Q}\mathbf{x}(K)$ is a adratic form in $\mathbf{x}(K-1)$ and $\mathbf{m}(K-1)$. us, J_0 can be written as

$$= \mathbf{x}'(K-1)\mathbf{P}\mathbf{x}(K-1) + \mathbf{x}'(K-1)\times \mathbf{S}\mathbf{m}(K-1) + \mathbf{m}'(K-1)\mathbf{R}\mathbf{m}(K-1) \tag{12}$$

ere $\mathbf{R}$ and $\mathbf{P}$ are symmetric matrices. ce $\mathbf{m}(K-1) = \mathbf{m}^\circ(K-1)$ minimizes

$$\Big|_{\mathbf{m}=\mathbf{m}^\circ} = 0 = \mathbf{x}'(K-1)\mathbf{s}_i + 2\mathbf{m}^{\circ\prime}\mathbf{r}_i \tag{13}$$

ere $\mathbf{s}_i$ and $\mathbf{r}_i$ are vectors whose com- ents are the components of the column of $\mathbf{S}$ and $\mathbf{R}$, respectively. m equation 13

$$(K-1)\mathbf{S} = -2\mathbf{m}^{\circ\prime}\mathbf{R} \tag{14}$$

is,

$$f_0 = \mathbf{m}'(K-1)\mathbf{R}\mathbf{m}(K-1) - \mathbf{m}^{\circ\prime}(K-1)\mathbf{R}\mathbf{m}^\circ(K-1) - 2\mathbf{m}^{\circ\prime}(K-1)\mathbf{R}[\mathbf{m}(K-1) - \mathbf{m}^\circ(K-1)] = [\mathbf{m}'(K-1) - \mathbf{m}^{\circ\prime}(K-1)]\mathbf{R}\times [\mathbf{m}(K-1) - \mathbf{m}^\circ(K-1)] \tag{15}$$

v, if a white gaussian noise dis- bance $\mathbf{u}(K-1)$ is added to the plant, is a quadratic form in $\mathbf{m}(K-1)$, $-1)$, and $\mathbf{u}(K-1)$.

$$\mathbf{x}'\mathbf{P}\mathbf{x} + \mathbf{x}'\mathbf{S}\mathbf{m} + \mathbf{m}'\mathbf{R}\mathbf{m} + \mathbf{u}'\mathbf{F}\mathbf{u} + \mathbf{u}'\mathbf{G}\mathbf{m} + \mathbf{u}'\mathbf{D}\mathbf{x} \tag{16}$$

re the arguments have been dropped porarily

$$= E[\mathbf{x}'\mathbf{P}\mathbf{x} + \mathbf{x}'\mathbf{S}\mathbf{m} + \mathbf{m}'\mathbf{R}\mathbf{m}] + E[\mathbf{u}'\mathbf{F}\mathbf{u}] \tag{17}$$

the other terms have dropped out use $\mathbf{u}(K-1)$ is *white* and, hence, rrelated with $\mathbf{x}(K-1)$ and $\mathbf{m}(K-1)$. n

$$-f_0 = E\{[(\mathbf{m}'(K-1) - \mathbf{m}^{\circ\prime}(K-1)]\times \mathbf{R}[\mathbf{m}(K-1) - \mathbf{m}^\circ(K-1)]\} \tag{18}$$

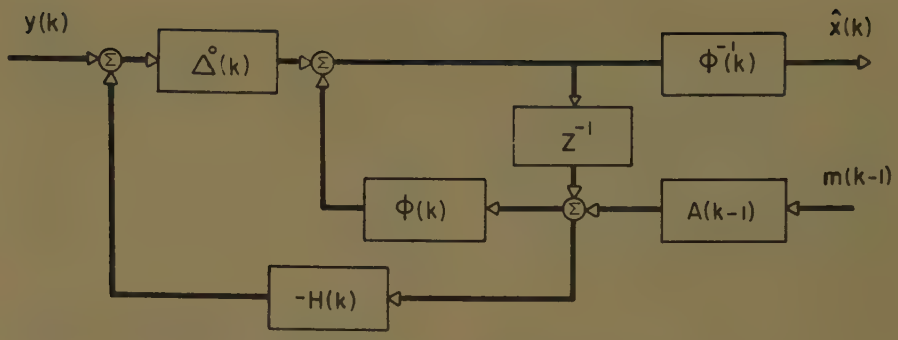


Fig. 2. Optimum estimator when $\mathbf{m} \neq 0$

Now $m(K-1) = \hat{m}^\circ(K-1)$ is shown to minimize the foregoing expression and, therefore, J_0.

First: Since f_0 is the absolute minimum value of J_0, $J_0 - f_0 > 0$ whenever $\mathbf{m}(K-1) \neq \mathbf{m}^\circ(K-1)$. Thus, from equation 15 it is obvious that $\mathbf{R}$ is a positive definite matrix.

Second: Let $\mathbf{m}^\circ(K-1) = \hat{\mathbf{m}}^\circ(K-1) + \tilde{\mathbf{m}}^\circ(K-1)$

$$E[J_0 - f_0] = E[(\mathbf{m}' - \hat{\mathbf{m}}^{\circ\prime} - \tilde{\mathbf{m}}^{\circ\prime})\times \mathbf{R}(\mathbf{m} - \hat{\mathbf{m}}^\circ - \tilde{\mathbf{m}}^\circ)] = E[\tilde{\mathbf{m}}^{\circ\prime}\mathbf{R}\tilde{\mathbf{m}}^\circ + (\mathbf{m}' - \hat{\mathbf{m}}^{\circ\prime})\times \mathbf{R}(\mathbf{m} - \hat{\mathbf{m}}^\circ) - 2\tilde{\mathbf{m}}^{\circ\prime}\mathbf{R}(\mathbf{m} - \hat{\mathbf{m}}^\circ)] \tag{19}$$

Now, $\tilde{\mathbf{m}}^\circ(K-1)$ is a linear combination of the $\tilde{\mathbf{x}}(K-1)$ and $\hat{\mathbf{m}}^\circ(K-1)$ is in $Y(K-1)$. Thus, by property 1

$$E[\tilde{\mathbf{m}}^{\circ\prime}\mathbf{R}\hat{\mathbf{m}}^\circ] = 0 \tag{20}$$

Also for $\mathbf{m}(K-1)$ to be realizable, it must be in $Y(K-1)$. Hence,

$$E[\tilde{\mathbf{m}}^{\circ\prime}\mathbf{R}\mathbf{m}] = 0 \tag{21}$$

and

$$E[J_0 - f_0] = E[\tilde{\mathbf{m}}^{\circ\prime}\mathbf{R}\tilde{\mathbf{m}}^\circ + (\mathbf{m}' - \hat{\mathbf{m}}^{\circ\prime})\mathbf{R}(\mathbf{m} - \hat{\mathbf{m}}^\circ)] \tag{22}$$

Since the first term on the right is independent of $\mathbf{m}$, and $\mathbf{R}$ is positive definite, the value of $\mathbf{m}$ which minimizes $E[J_0 - f_0]$ is given by

$$\mathbf{m}(K-1) = \hat{\mathbf{m}}^\circ(K-1) \tag{23}$$

By the principle of optimality, this is the value of $m(K-1)$ which minimizes I_K.

To continue

$$J_1 = \mathbf{x}'(K-1)\mathbf{Q}\mathbf{x}(K-1) + J_0 = \mathbf{x}'(K-1)\mathbf{Q}\mathbf{x}(K-1) + f_0 + [\mathbf{m}'(K-1) - \mathbf{m}^{\circ\prime}(K-1)]\mathbf{R}[\mathbf{m}(K-1) - \mathbf{m}^\circ(K-1)] \tag{24}$$

In the absence of disturbances, the first two terms on the right are quadratic forms of $\mathbf{x}(K-2)$ and $\mathbf{m}(K-2)$ and, therefore, in complete analogy with the preceding work.

$$E[J_1 - f_1] = E\{[\mathbf{m}'(K-2) - \mathbf{m}^{\circ\prime}(K-2)]\times \mathbf{R}_1[\mathbf{m}(K-2) - \mathbf{m}^\circ(K-2)]\} + E\{[\mathbf{m}'(K-1) - \mathbf{m}^{\circ\prime}(K-1)]\times \mathbf{R}[(\mathbf{m}(K-1) - \mathbf{m}^\circ(K-1)]\} \tag{25}$$

By the principle of optimality, the minimum value of $E[J_1 - f_1]$ is

$$\min_{m(K-2)} \{E[\mathbf{m}'(K-2) - \mathbf{m}^{\circ\prime}(K-2)]\times \mathbf{R}_1[\mathbf{m}(K-2) - \mathbf{m}^\circ(K-2)] + E\tilde{\mathbf{m}}^{\circ\prime}(K-1)\mathbf{R}\tilde{\mathbf{m}}^\circ(K-1)\} \tag{26}$$

Note that by property 5, $\mathbf{m}^\circ(K-1)$ does not depend upon $\mathbf{m}(K-2)$. Thus, the optimum value of $\mathbf{m}(K-2)$ is the one which minimizes $E[\mathbf{m}'(K-2) - \mathbf{m}^{\circ\prime}(K-2)]\mathbf{R}_1[\mathbf{m}(K-2) - \mathbf{m}^\circ(K-2)]$. Hence, the optimum $\mathbf{m}(K-2)$ is given by

$$\mathbf{m}(K-2) = \hat{\mathbf{m}}^\circ(K-2) \tag{27}$$

Clearly, there is now proof by induction that the optimum realizable $\mathbf{m}(k)$ is given by

$$\mathbf{m}(k) = \hat{\mathbf{m}}^\circ(k) \tag{28}$$

This completes the proof.

Remarks

1. The theorem allows the design of a control system to be broken into two independent design problems: (1) the determination of the matrix $\mathbf{B}^\circ(k)$, which is a problem that is solved in reference 1, and (2) the determination of a system to generate $\hat{\mathbf{x}}(k)$. Such a system is shown in Fig. 2 where only the matrix $\mathbf{A}^\circ(k)$ is unknown. The problem of determining $\mathbf{A}^\circ(k)$ is solved in reference 8. Thus, a complete design procedure is available.

2. The theorem could be extended to include performance indices of the form

$$I_K = \sum_{k=1}^{K} [\mathbf{x}'(k)\mathbf{Q}(k)\mathbf{x}(k) + \mathbf{m}'(k)\mathbf{P}(k)\mathbf{m}(k)] \tag{29}$$

3. By property 2, $\hat{\mathbf{m}}^\circ(k)$ is the optimum linear-control signal if the disturbance is nongaussian.

4. Nonwhite inputs and disturbances can be handled by considering them to be the outputs of a linear dynamic system excited by white noise.[8,9] The control system is then designed by considering an *augmented plant*, consisting of the original plant plus the fictitious linear dynamic systems, which generate the noise. Nonregulator problems can be handled by a similar technique.

Conclusions

If the controller and the estimator are independently optimized, an optimum control system, with respect to a quadratic performance criterion, results. Since techniques are already in existence for optimizing controller and estimator, in single and multivariable, stationary and nonstationary plants, a systematic general method of designing linear discrete data control systems, which are subject to random inputs and disturbances, is available.

References

1. Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indexes, **R. E. Kalman, R. W. Koepcke.** *ASME Transactions*, American Society of Mechanical Engineers, New York, N. Y., 1958.

2. On the General Theory of Control, **R. E. Kalman.** *"Proceedings*, International Federation of Automatic Control Congress, 1960," Butterworth's Scientific Publications, London, England, 1961.

3. On Optimal Computer Control, **J. E. Bertram, P. E. Sararchik.** *Ibid.*

4. General Synthesis Procedure for C puter Control of Single-Loop and Multil Linear Systems, **R. E. Kalman, J. E. Bertra** *AIEE Transactions*, pt. II (*Applications Industry*), vol. 77, 1958 (Jan. 1959 section), 602–09.

5. A Method for the Symbolic Represen tion and Analysis of Linear Periodic Feedb Systems, **E. O. Gilbert.** *Ibid.*, vol. 78, 1959 (1960 section), pp. 512–23.

6. Dynamic Programming (book), **R. E. B man.** Princeton University Press, Princet N. J., 1957.

7. Use of a Mathematical Error Criter in the Design of Adaptive Control Syst **C. W. Merriam.** *AIEE Transactions*, pt. (*Applications and Industry*), vol. 78, 1959 (1960 section), pp. 506–12.

8. A New Approach to Linear Filtering Prediction Problems, **R. E. Kalman.** *Jou of Basic Engineering*, New York, N. Y., Mar. 1

9. A Simplified Derivation of Linear Le Square Smoothing and Prediction The **H. W. Bode, C. E. Shannon.** *Proceedings*, Insti of Radio Engineers, New York, N. Y., 1950.

Phase-Space Analysis and Design of Linear Discontinuously Damped Feedback Control Systems

K. W. HAN STUDENT MEMBER AIEE

G. J. THALER MEMBER AIEE

LINEAR SYSTEMS can be designed having very fast responses with essentially deadbeat performance by the inclusion of compensating loops which are switched in and out of the system as required. Fast response is obtained by using an unstable or nearly unstable uncompensated system to provide the desired rise time. Deadbeat performance is obtained by designing the compensation loops so that they provide an overdamped system when the compensation is switched in. Selection of the real roots for the overdamped system is based on the desired location of the eigenvectors in the phase space. A switching computer is required which connects the compensation loops as the state point reaches a hyperplane which is related to the eigenvector in a special way. The computer is readily realized from derivative signals and considerable engineering simplification is permissible in both the compensation design and the switching computer because of the topological nature of the phase portrait in the vicinity of properly chosen eigenvectors. Experimental results verify the theoretical conclusions.

Nomenclature

ζ = damping factor
E = error
θ_R = input quantity
θ_c = output quantity
N, n = integers
A, a = coefficients
r = root of characteristic equation
τ = pole of open-loop function
s = Laplace operator
P_s = location of the state point in the phase space at the switching instant
$G(s)$ = open-loop transfer function in the forward path
$H(s)$ = open-loop transfer function in a feedback path
K = forward gain of open-loop system

Theoretical Background

It is clear from linear theory that a feedback control system with high gain has fast response and is accurate in both static and steady states. Unfortunately, high gain is usually accompanied by light damping with consequent oscillatory transient response. Various investigators have proposed schemes which essentially vary the damping as some function of the error to obtain small ζ when the error is large and vice versa. Perhaps the s plest and most obvious approach is t of varying the gain of the main ampli as suggested by Blumenthal and Be Another early and attractive propos that of Lewis,[2] which uses a nonlinear nal to produce a smooth variation throughout the range of operating inp Other proposals have suggested a continuous change in ζ. Two such vestigations have dealt with relay serv where the damping loop was permane connected but was operative only in relay dead zone. Still others have concerned with switching in a feed loop. Flügge-Lotz[5] and Taylor pro to alter both position and velocity f back in step fashion according to a determined schedule which is implemen by a fairly complex computer. Meik proposes switching in main feedback from positive feedback for fast rise t to negative feedback for good damp For second-order systems switching a an eigenvector is easily instrumented, gain design suitable for overdam dynamic response may not be compat with static accuracy requirements. third-order systems eigenvector (f mathematical definition see referenc switching is recommended. (In term phase-space trajectories it may be s A linear differential equation has only singular point; for each real root of

Paper 61-828, recommended by the AIEE back Control Systems Committee and app by the AIEE Technical Operations Departme presentation at the AIEE Summer General ing, Ithaca, N. Y., June 18–23, 1961. Manu submitted May 3, 1960; made availabl printing April 25, 1961.

K. W. Han and G. J. Thaler are both o U. S. Naval Postgraduate School, Monterey

fferential equation there is one eigen-ctor; each eigenvector appears in e phase space as a phase trajectory hich is a straight line through the sin-lar point.) The mathematical develop-ent of the switching criterion follows e matrix methods of Bogner and Kazda[8] d leads to a rather complex switching mputer, with two switching operations normal operation, though simplifica-n of the computer and use of only one itching operation applies under special nditions. Another discontinuous damp-g scheme is proposed by Ostrovskii,[9] ho suggests switching in feedback paths alter the coefficients of the closed loop fferential equation. The results reported this paper, with advantages as indicated, e an extension of this latter concept.

hase-Space Analysis

The differential equation of an Nth der linear feedback control system may expressed as

$$E+A_1\dot{E}+A_2\ddot{E}+\ \ldots A_{n-1}E^{n-1}+A_nE^n = f(\dot{\theta}_R, \ddot{\theta}_R \ldots) \quad (1)$$

e characteristic equation has n roots, hich may be real, imaginary, or complex pending on the values of the A's. agine two sets of values for the co-icients, where a dominant pair of roots deliberately set near the imaginary axis provide a suitable rise time for one set d all roots are real and negative for the ond set. Since both equations are of e same order their phase portraits may studied in the same phase space. For e underdamped case the singular point a focus and the phase trajectories con-rge on this focus. When the system is ble all trajectories ultimately reach e origin of the error phase space provid-:

The system is type 1 and is excited by tial conditions and/or a step displace-nt.

The system is type 2 and is excited by tial conditions and/or a step displace-nt and/or a ramp function, etc.

r the overdamped system (all real nega-e roots) the entire phase space is filled h trajectories which terminate at the gin (for the previously mentioned con-ions) and most of these trajectories hibit a monotonic variation. It is well own that for each of the real roots there n eigenvector, which corresponds to a se trajectory that is a straight line. It possible to construct hyperplanes so t each hyperplane corresponds to an envector and also contains a subset of complete phase trajectories. When the hyperplanes containing a complete subset of phase trajectories are chosen they subdivide the phase space into regions so that no phase trajectory can pass out of one region and into another. Thus the hyperplanes act as boundaries which funnel the phase trajectories into the origin.

During the operation of the system only one of the two equations can be effective at a given instant, but for purposes of analysis it is assumed that both families of phase trajectories exist simultaneously. At every point in the space, two trajectories, and only two, intersect; one curve is from each family. If a step displacement is applied to the underdamped system the state point follows a selected trajectory which intersects an infinite number of phase trajectories of the overdamped system. If the proper compensation circuits have been devised to change the coefficients of the characteristic equation from the selected underdamped case to the selected overdamped case, then the switch inserting these may be thrown at any time and the state point transfers smoothly from the underdamped trajectory to the intersecting overdamped trajectory. To prove this it is only necessary to note that the coordinates of the state point at the switching instant are

$$P_s=(E_s, \dot{E}_s, \ddot{E}_s \ldots E_s^{n-1})$$

Now this point lies on both phase trajectories and thus satisfies both characteristic equations. All derivations are continuous except the nth, in which discontinuity is permissible. Note that this condition applies for any arbitrarily selected switching point, and as a result a transientless switching operation is always obtained.

To provide automatic switching at the preselected point, note that a straight line through the origin of the phase space and through the selected switching point has direction numbers

$$E_s, \dot{E}_s, \ddot{E}_s, \ldots E_s^{n-1}$$

and the symmetric equations of this line are

$$\frac{E}{E_s}=\frac{\dot{E}}{\dot{E}_s}=\frac{\ddot{E}}{\ddot{E}_s}=\frac{E^{n-1}}{E_s^{n-1}} \quad (2)$$

This is not a convenient form for implementation of a switching computer, so it is noted that hyperplane containing this line is given by

$$a_0E+a_1\dot{E}+a_2\ddot{E}+\ldots a_{n-1}E^{n-1}=0 \quad (3)$$

where the coefficients a_0, a_1, etc., are chosen so that

$$a_0E_s+a_1\dot{E}_s+a_2\ddot{E}_s+\ldots a_{n-1}E_s^{n-1}=0 \quad (4)$$

Equation 3 indicates that the only switching computer needed is an adding amplifier, provided that all of the derivative signals are available or can be generated. Such a switching computer is sensitive to all of the points on the hyperplane defined by equation 3 and it is necessary that this hyperplane be carefully chosen if specified performance is to be obtained. When a step displacement input is the only input to be considered only those phase trajectories starting on the error axis are important. Assume that a switching point on one of these step response trajectories is chosen to define equations 2, 3, and 4. All other step trajectories pierce the hyperplane of equation 3 along a straight line through the origin which may be considered a mapping of the E-axis in the hyperplane. Thus, for step displacement inputs, only a line in the hyperplane is actually used for switching and any other hyperplane which contains this line may be instrumented and used if more convenient. Note that while the step response is insensitive to the hyperplane chosen, the response to other inputs is affected. It should also be noted that a computer designed to implement equation 3 produces an output of which the sign (or polarity) depends on the sign or direction of the step input. Then for use with a polarity sensitive relay a satisfactory switching equation is

$$\dot{E}(a_0E+a_1\dot{E}+a_2\ddot{E}+\ldots a_{n-1}E^{n-1})=0 \quad (5)$$

Several practical modifications of this are discussed in a later section.

The system is discontinuously damped because the operation of the switch introduces circuits which change the roots from complex values to real, negative values. Nevertheless the system may be considered linear for step displacement inputs. Note that the switching hyperplane subdivides the phase space into two regions, and in each region the system is completely linear but in each region a different linear differential equation applies. However, the switching is a straight line through the origin of the phase space, and this insures that the initial conditions after switching are always directly proportional to the magnitude of the step. This leads to the following characteristics:

1. The rise time to the switching point is the same for all magnitudes of step.

2. The settling time of the composite system is constant.

3. The per-cent overshoot, if any, is constant.

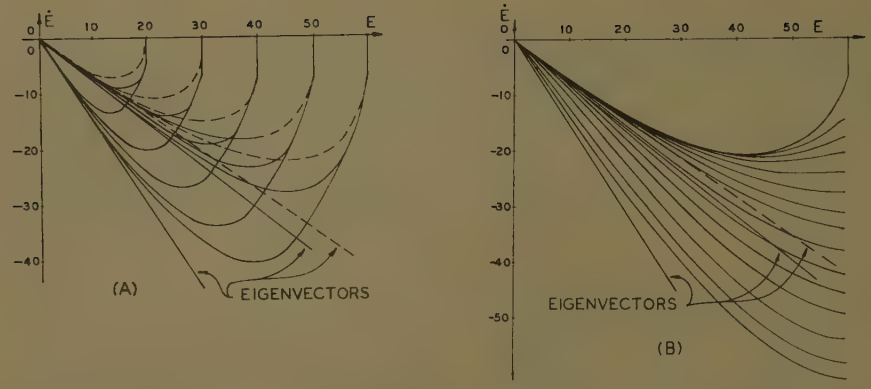


Fig. 1. Phase plots of a third-order system with discontinuously damped feedback

$\dddot{\theta}_c+2.20\ddot{\theta}_c+0.54\dot{\theta}_c+\theta_c=\theta_R$ (underdamped)

$\dddot{\theta}_c+3\ 5\ddot{\theta}_c+3.5\dot{\theta}_c+\theta_c=\theta_R$ (overdamped)

A—Switching in the hyperplane

B—Switching not in the hyperplane

These features are independent of the choice of a switching line, but the numerical values associated with each depend very much on this choice. Since the purpose of discontinuous damping is to permit a very fast rise time with deadbeat (or nearly deadbeat) response and a minimum settling time, the following considerations are important: For fast rise time switching should be delayed until E is as near to zero as is permissible; for deadbeat response due to the heavily damped system, a trajectory must be selected which is nearly a straight line to the origin (i.e., a fast eigenvector if this is possible) since departure from this choice leads to the long-tailed response (long settling time) characteristics of overdamped systems.

For second-order systems switching on the eigenvector is possible and practical. For higher order systems the trajectory of the underdamped system selected by the step displacement input does not pass through that line in the phase space which is the desired eigenvector and therefore the optimum trajectory is not available. However, for each eigenvector there is a hyperplane which has a subset of trajectories lying entirely in it and these trajectories become tangent to the eigenvector at the origin. In general the equation of this hyperplane is known to be one order lower than the characteristic equation and is formulated by removing the root corresponding to the eigenvector. Thus,

$$E^{n-1}+\sum_{i=2}^{i=n} r_i E^{n-2}+\ldots+\prod_{i=2}^{i=n} r_i E=0 \qquad (6)$$

is the equation for the hyperplane for removed r_1. It is also the defining equation for the switching computer needed to introduce the discontinuous damping.

Fig. 1 shows two phase portraits on the E versus $\dot{E}$ plane for an overdamped third-order system. Note that in both cases the three eigenvectors can be located, and are straight lines as expected. Fig. 1(A) was obtained so that all of the trajectories around the eigenvectors lie only in the hyperplanes associated with that eigenvector. For Fig. 1(B) the initial conditions were chosen so that the trajectories are in the vicinity of, but not actually in, the hyperplane.

Theoretical Considerations

In the design of a discontinuously damped system of this type, three theoretical, (and other practical) problems arise, such as the design of the uncompensated and compensated linear systems, the selection of the hyperplane, and the design of the switching computer. Generally the uncompensated system is of single-loop design with gain set to satisfy both the steady-state accuracy requirements and the rise time requirements. The resulting system is expected to be very lightly damped, perhaps unstable. As far as theory is concerned an unstable system is permissible; in practice however, a stable system is more desirable because of the possibility of a failure in the switching loop, so some fixed compensation may be used. The design of the overdamped system is accomplished very simply by arbitrarily selecting a suitable set of real roots and evaluating the coefficients of the corresponding characteristic equations, and then computing the gain constant for the various derivative signal channel which must be used as compensators. When the input signal is restricted to a step displacement input, the derivatives of error are numerically equal to the derivatives of the output signal except for a factor of -1.0. The generalized block diagram is as shown in Fig. 2. The equations (assuming that $G(s)$ has no zeros) are

$$G(s)=\frac{K}{s^n+(\tau_1+\tau_2+\ldots)s^{n-1}+\ldots+(\tau_1\tau_2\tau_3\ldots)s^x} \qquad (7$$

and the composite feedback function when the switch is closed is

$$H(s)=1+As+Bs^2+\ldots+Ms^N \qquad (8$$

where $n>N$, and normally $n=N+$ Then

$$G(s)H(s)=\frac{K(1+As+Bs^2+\ldots+Ms^N)}{s^n+(\tau_1+\tau_2+\ldots)s^{n-1}+\ldots+(\tau_1\tau_2\tau_3\ldots)s^x} \qquad (9$$

and if $n=N+1$ the characteristic equation becomes

$$0=s^n+(\tau_1+\tau_2+\ldots+KM)s^{n-1}+\ldots(\tau_1\tau_2\tau_3+KX)s^x+\ldots KBs^2+KAs+K \qquad (1$$

For a type 1 system $X=1$ and the characteristic equation becomes

$$0=s^n+(\tau_1+\tau_2+\ldots+KM)s^{n-1}+\ldots+(\tau_1\tau_2\tau_3\ldots+KA)s+K \qquad (1$$

All τ's are known from the uncompensated system; K is known; and each desired coefficient is known from the arbitrary selection of the real roots. Thus, each coefficient in equation 11 may be equated to the corresponding coefficient for the desired overdamped system and A, B, $\ldots$ N are evaluated. The choice of real roots is based on a desire to obtain rapid damping of the response during the overdamped mode of operation. To obtain this, one, and only one, of the real roots

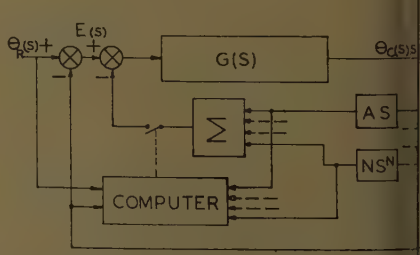


Fig. 2. Generalized block diagram for r order system with discontinuous feedb compensation

osen is quite small; all others chosen e quite large. This combination gives e to an eigenvector, which projects on the E versus $\dot{E}$ plane at a relatively all angle to the $\dot{E}$-axis. The hyperne associated with this eigenvector used as the switching surface. When is is done the initial conditions at the itching instant cause the residue at the allest root to go to zero. Thus, the tling time is controlled solely by the ge roots and as a consequence is quite all. To show that the smallest real t does not appear in the transient ponse after switching, consider a rd order system for which the differtial equation is

$$\dddot{E}+(r_1+r_2+r_3)\ddot{E}+(r_1r_2+r_1r_3+r_2r_3)\dot{E}+r_1r_2r_3E=0 \quad (12)$$

d the real roots are $r_1<r_2<r_3$.
Equation 12 has a solution

$$E)=\beta_1e^{-r_1t}+\beta_2e^{-r_2t}+\beta_3e^{-r_3t} \quad (13)$$

ere

$$=\frac{r_2r_3E_s+(r_2+r_3)\dot{E}_s+\ddot{E}_s}{(r_2-r_1)(r_3-r_1)} \quad (14)$$

$\dot{E}s$, and $\ddot{E}s$ are the initial conditions d β_2 and β_3 are not relevant to this cussion. Now the hyperplane at ich switching occurs is chosen to be

$$E+(r_2+r_3)\dot{E}+\ddot{E}=0 \quad (15)$$

erefore, when switching occurs at any nt in the hyperplane, the initial condins at that point force the numerator of uation 14 to zero, and the $\beta_1e^{-r_1t}$ term appears from equation 13. This is e for any order equation. For a thirder system such switching reduces the onse to a second-order response. us, motion is in the hyperplane. For her order systems motion is confined a subspace of order M-1. Since every nt on the resulting phase trajectory sfies equation 6, the switching comer output remains at zero and the tch has no tendency to reopen.

he function of the computer is to opte the switch when the state point hes the selected hyperplane. Thus, ation 6 must be mechanized, which is tively simple since it may be assumed t all derivative signals have been made ilable to provide the compensation . All that remains in the scaling of magnitudes of these signals to provide coefficients defined by the chosen roots the overdamped system. This can ally be done with potential dividers. n the signals are totaled to formulate ation 6 and the summation signal is fed into a circuit which operates a relay when the signal goes to zero. One way for the polarity-sensitive relay circuit to recognize the difference between positive and negative steps is to multiply equation 6 by $\dot{E}$. This complicates the computer mechanization, however, and for many applications other methods may be substituted. For an nth order system, equation 6 is the equation of the hyperplane, which corresponds to the eigenvector, and thus is the defining equation for the switching computer. If this hyperplane is used as a switching surface for step displacement inputs, then each phase trajectory starts on the E-axis and pierces the hyperplane at some point which may be called a switching point. The loci of these switching points form a straight line on the surface of the hyperplane which passes through the origin. Only the points on this line are used in switching. Therefore, if any other hyperplane can be instrumented in the switching computer so that the new hyperplane intersects the switching hyperplane along a trace which is exactly the switching line, then the operation of the new switching hyperplane is exactly correct for step displacement inputs. Since the switching line is a straight line through the origin, its projection onto the E versus $\dot{E}$ plane is also a straight line through the origin. The equation of this line is readily computed and is of the form

$$E+A\dot{E}=0 \quad (16)$$

A switching computer to instrument this is a very simple adder, and because of the simplicity the coefficient A may be set by adjustment rather than calculation. The surface thus defined is a hyperplane which (in the 3-dimensional case) is perpendicular to the $\dot{E}$ versus E plane. When switching occurs in response to a step input, the state point does not remain in the surface defined by equation 16, but in the surface defined by equation 6. Thus, the switching computer develops an output and the relay will reoperate unless proper precautions are taken in the design of the switching computer.

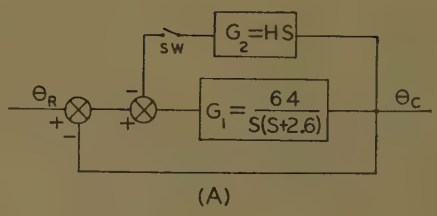
Analytic and Computer Verification

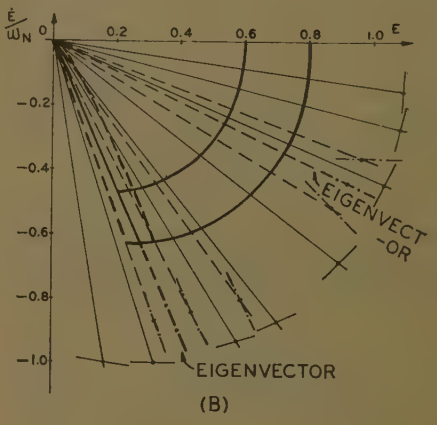
Second-Order System

Consider the block diagram of Fig. 3(A) with transfer function as given in the blocks. The differential equation of the system without compensation is

$$\ddot{E}+2.6\dot{E}+64E=0 \quad (17)$$

for which





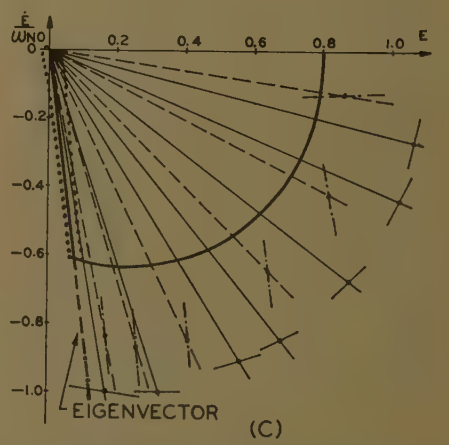


Fig. 3. Block diagram and phase-plane plots of a second-order system with discontinuous tachometer feedback

A—Block diagram
B—With moderate tachometer feedback
C—With large tachometer feedback

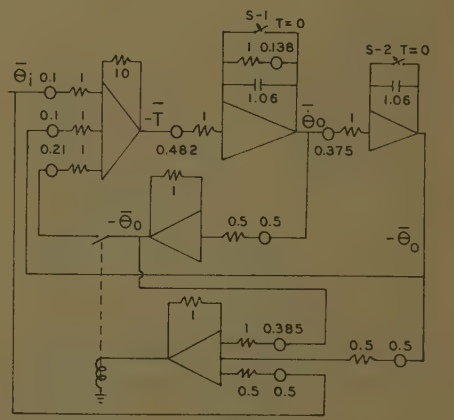


Fig. 4. Analog computer set up for a second-order system with discontinuous tachometer feedback

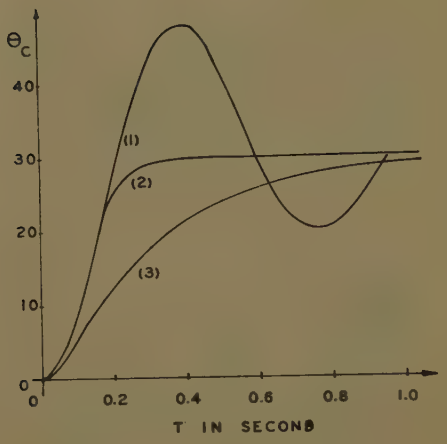


Fig. 5(A). Analog computer plots for a second-order system with discontinuous tachometer feedback

1—underdamped
2—discontinuously damped
3—overdamped

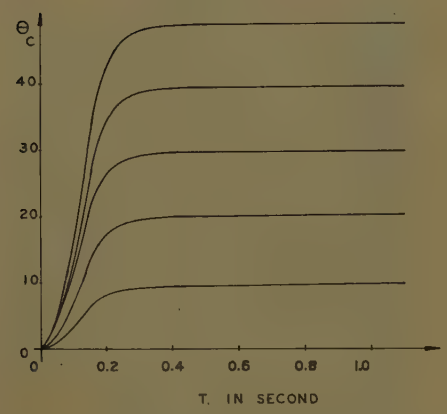


Fig. 5(B). Transient response for various magnitudes of step input

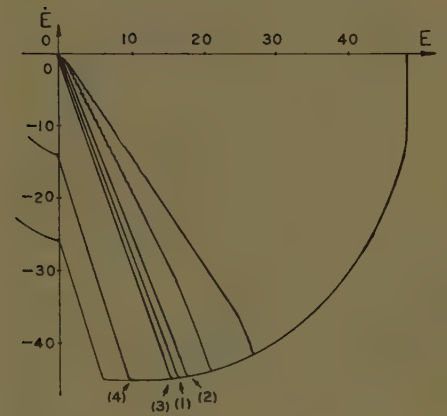


Fig. 5(C). Illustration of the effects of variations in the switching computer adjustment

1—Switching on the eigenvector
2—Switching early
3—Switching later
4—Switching too late

$\zeta = 0.162$, $\omega_n = 8$

For the compensated system with $H=0.3$ the equation is

$$\ddot{E}+21.8\dot{E}+64=0 \tag{18}$$

for which the real roots are at -18.3 and -3.5. For $H=0.96$

$$\ddot{E}+64\dot{E}+64=0 \tag{19}$$

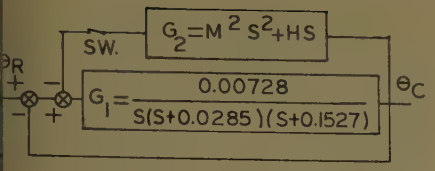
and the real roots are at -0.995 and -63.0. For the second-order system the eigenvectors of the overdamped equations are in the E versus $\dot{E}$ plane, and thus can be used as switching lines. The two eigenvectors can be located easily since they are precisely those isoclines for which the slope of the isocline is identical with the slope of the trajectory and also is the value of the real root. To illustrate this, Fig. 3(B) shows an isocline plot for equation 18, and Fig. 3(C) shows an isocline plot for equation 19. Typical trajectories for operation as a discontinuous system have been constructed and it is apparent that slight inaccuracies in switching do not cause a significant difference.

This system was simulated in the analog computer as shown on Fig. 4. The switching line is instrumented very simply as a summer amplifier operating a normally closed 2-position relay. When the step is applied the relay automatically opens and remains open until the eigenvector is reached, at this point the relay voltage reduces to zero and the relay drops out, closing the damping circuit. For all points on the eigenvector this voltage sum remains zero and the relay is not actuated. The relay is sensitive only to magnitudes, so no additional devices are required to distinguish between positive and negative steps. Fig. 5(A) compares the step responses of the underdamped, overdamped, and discontinuously damped systems. Fig. 5(B) shows the transient responses obtained with various magnitudes of step input, and Fig. 5(C) shows the effect of variations in the switching computer adjustment. The slope of the switching line is readily adjusted by changing the magnitude of the velocity signal fed to the computer. Fig. 5(C) indicates that a wide range of adjustment is permissible without significant changes in the response. This, however, is due more to the characteristics of the switch and computer than to the phase-plane topology. Curve *1* in Fig. 5(C) is the eigenvector trajectory and the system operates as predicted, with no additional operations of the relay. For earlier switching the trajectory should seek the slow eigenvector and thus should be concave upward. Actually the de viation of the trajectory from the switch ing line caused the relay to reopen re peatedly so that the trajectories show are due to the system chattering dow the switching line. For late switching the trajectories, in this case, are so nearl straight that the relay did not reope ate until the switching delay produce curve *4* and reversion to the underdampe condition then caused a tremendous ove shoot. Thus, this particular choice of switching scheme permits early switchin over as wide a range as desired, provid that relay chatter is acceptable. Th actual trajectory obtained with rela chatter provides a faster response tha would be obtained if it were possible follow the overdamped trajectory fro the same initial switching point. La switching is obviously dangerous in th case. The width of the permissive zo for late switching depends on the sensiti ity of the relay-switching computer un and if the gain of this circuit is high, la switching may not be acceptable at because of the possibility of reoperati of the relay, which would cause a lar overshoot.

Third-Order System

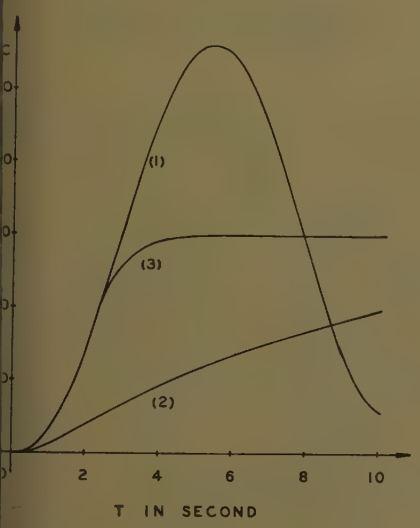
The block diagram of Fig. 6(A) sho the system and transfer functions. Th roots of the uncompensated system a at -0.180, and $-0.0006+j0.0625$. Fi the compensated case the roots we chosen at -0.0115, -0.178, and -0.3 from which the gains required in the fee back path are $H=44.5$ and $M=15$ Fig. 6(B) shows the analog computer i plementation including the switching c cuit which once more is just a summ amplifier driving a normally closed rela Note that a normal differentiator-amp fier and a simple R-C circuit were used obtain the required $\ddot{\theta}_0$ signal. This w done because practical applications no mally require the simplicity of the M approximation, and therefore it is i portant to know whether the added p disturbs either the damping or the switc ing. Also, a performance comparis between the exact and approximate s tems seemed necessary.

Fig. 7(A) compares the transient sponses of the overdamped, underdamp and discontinuously damped systems, a Fig. 7(B) compares the frequency sponses. In Fig. 7(B) curve *1* is frequency response of the underdamp system, and also of the discontinuo system when the relay of switch negligible dead zone. (With sinusoi input, all steady-state signals vary peri ically and equation 6 is satisfied



ig. 6(A). Block diagram of third-order system

nfinitesimal time on each cycle. So ne damping is inserted for essentially ero time.) When the relay has a small ut appreciable dead zone, the discon-nuous damping is quite effective and the requency response is curve *2* for a wide ange of signal amplitudes. In the spe-fic case tested for a range of test signal mplitudes of 0–50 volts (above this the mplifiers saturated) curve *2* was ob-ained for all amplitudes from 7 to 35 olts. For amplitudes between 35 and 0 volts the system became underdamped nd approached curve *1*. For ampli-udes from 7 to 0 volts the frequency re-ponse approached the overdamped condi-on of curve *3*. In taking the data many utput waveshapes were distorted, and ne amplitude of the fundamental was stimated. The *R-C* differentiation was ot used for these curves. The phase ane is used in Fig. 8 to show the effect f the *R-C* differentiator on the system sponse. It is readily seen that the erformance using the *RC* circuit is negli-bly different from that using the theo-tically exact relationships. Early and te switching can also be accomplished this case merely by adjusting the co-ficient of the $\dot{E}$ term in the switching



g. 7(A). Third-order system transient response for a step input

1—Underdamped
2—Overdamped
3—Discontinuously damped

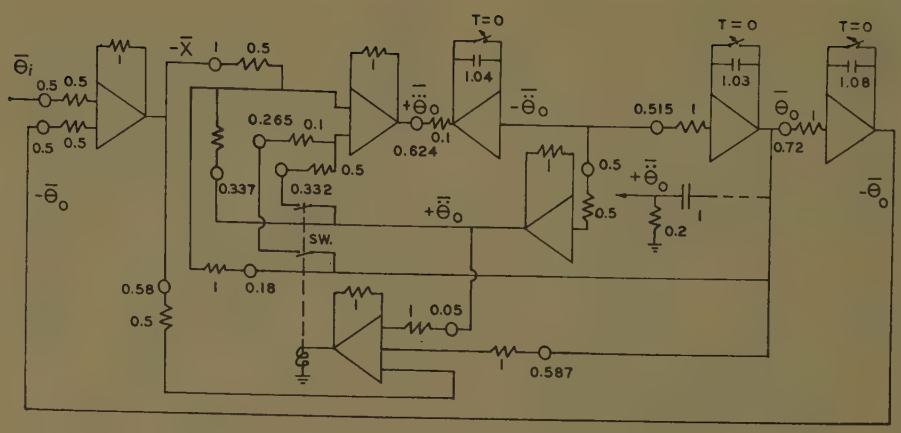


Fig. 6(B). Analog computer setup for a third-order system

Scaling, $\alpha' t = 0.09$, $\alpha E = \alpha\theta_i = \alpha\theta_o = 0.03$, $\alpha\dot{\theta}_o = 0.02$, $\alpha\ddot{\theta}_o = 0.01$, $\alpha\dddot{\theta}_o = 0.06$, $\alpha u = 0.2$

equation. The results are essentially the same as for the second-order case (see Fig. 5(C)) and the same comments apply.

Fourth-Order System

To extend the theory to higher order systems, to verify its applicability to actual physical systems, and to investigate the need for (and the effect of) practical approximations to the mathematical requirements, a closed-loop positioning servomechanism was assembled. This consisted of a d-c amplifier, amplidyne generator, and a 1/4-horsepower shunt motor in cascade. Direct-current excited potentiometers were used for an error detector. The load was a large inertia and a d-c tachometer was attached. The gain was set to provide a stable, but badly underdamped system. Open-loop frequency response tests provided a transfer function

$$\frac{\theta_c}{E}(s) = \frac{23{,}774}{s(s+8)(s+11.55+j11.8)\times(s+11.55-j11.8)} \qquad (20)$$

From this the roots of the uncompensated

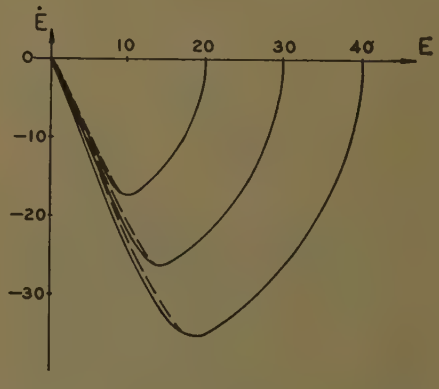


Fig. 8. Comparison of response using R-C differentiator (broken line) instead of computer differentiator (solid line)

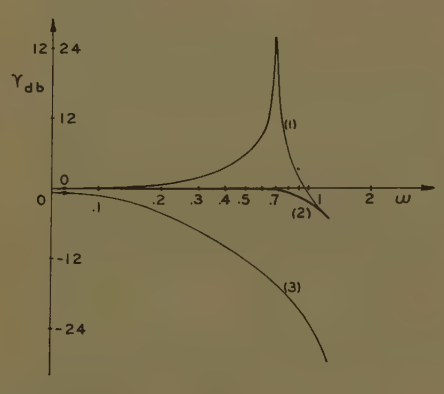


Fig. 7 (B). Frequency respose of a third-order system (computer setup)

1—Underdamped
2—Discontinuously damped
3—Overdamped

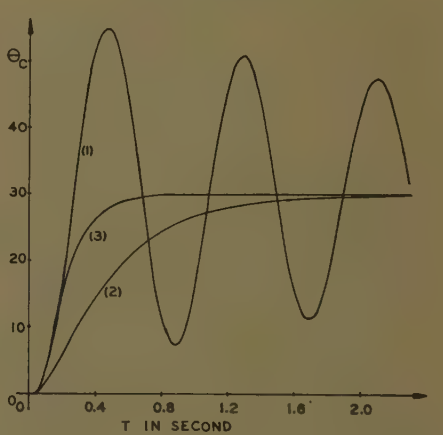


Fig. 9. Step response of a fourth-order system

1—Underdamped
2—Overdamped
3—Discontinuously damped

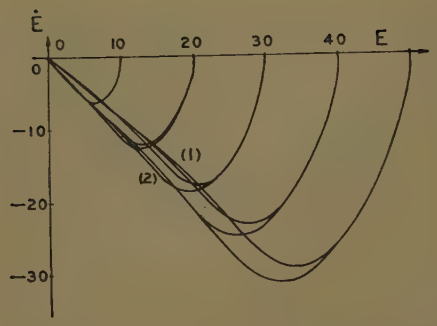


Fig. 10. Comparison of response

1—Using hyperplane switching
2—Using E versus Ė switching

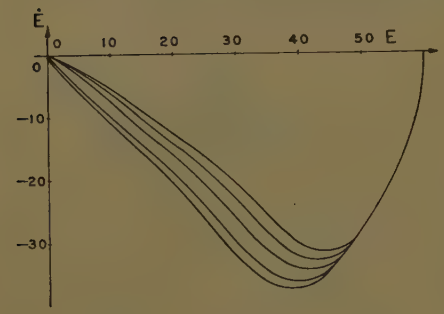


Fig. 11. Effect of switching time on response (computer study)

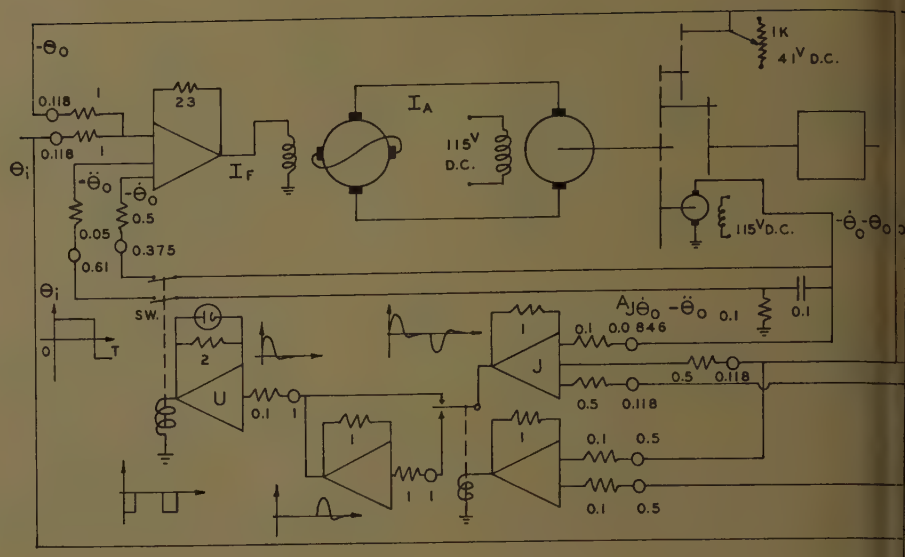


Fig. 12. Schematic diagram of the tested fourth-order system (Amplidyne, General Electric model 5AM79AB362, type AM; motor, General Electric, model 5BC, type BC)

system are at $-0.35+j8$ and $-15.15+j11.89$. During the tests RC differentiating filters were cascaded with the tachometer and various derivative signals observed. The second derivative signal was quite noisy and the third derivative signal was too noisy for use. Therefore it was decided not to use the third derivative signal at all, and preferably not the second derivative signal. However, it is easily shown that the system cannot be overdamped using tachometer feedback only, but using both first and second derivative feedback roots can be placed at -3, -6, and $-11.5\pm j33.7$. The system is completely overdamped for a step displacement input while complex conjugate roots exist. The second derivative signal was too noisy to use in the switching computer; thus, the computer had to operate on E and $\dot{E}$ signals only.

It was considered desirable to study the system on the analog computer first. The underdamped and overdamped systems were simulated with the indicated roots and the correct hyperplace switching surface was computed using the same scheme that applied to the second- and third-order cases. Typical transient response curves are shown by Fig. 9. A simplified switching equation was instrumented using E and $\dot{E}$ signals only in the form

$$E+A\dot{E}=0 \qquad (21)$$

The theory and calculation of coefficient A and some additional instrumentation requirements have been discussed. The E versus $\dot{E}$ plane is utilized in Fig. 10 to compare the results of switching at the theoretically correct hyperplane, with those obtained by switching at the hyperplane defined by equation 21. Fig. 11 indicates that the choice of the coefficient A is not critical.

Physical System Verification

Fig. 12 shows the schematic diagram of the control system tested. Parameter values are indicated. The modifications of the switching circuit are necessary to prevent reopening of the compensation loop. Fig. 13(A) shows a family of transient response curves obtained from the physical system, with range of step amplitudes in excess of 100 degrees. Fig. 13(B) shows the effect of varying the coefficient A in the switching equation. It is apparent that acceptable step response can be obtained over a reasonable range of adjustments of the switching computer.

Discussion and Conclusion

From the equations, circuits and computer studies presented it is apparent that the concept of discontinuous damping of linear control systems is a relatively simple concept: the required compensation loops are easily determined; the

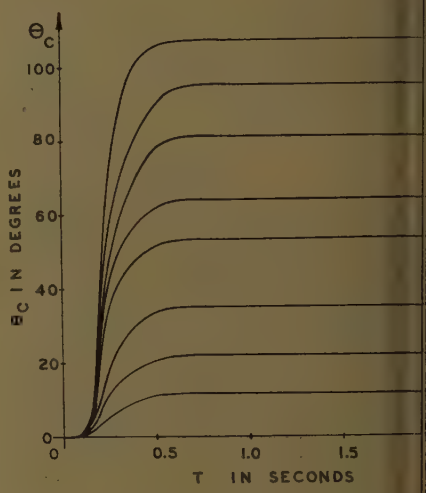


Fig. 13(A). Transient response for differ magnitudes of step input

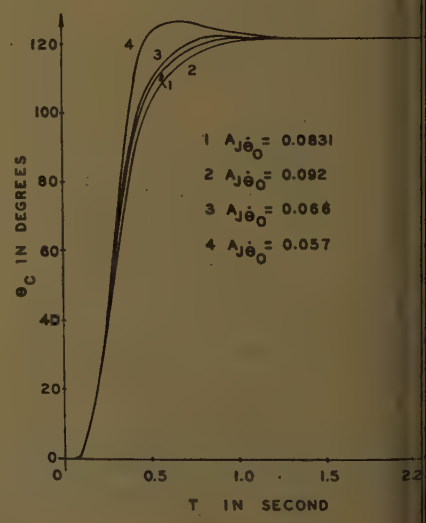


Fig. 13(B). Effects of switching time transient response (see Fig. 12 for

parameter values are easily calculated; and the switching computer for step inputs is readily designed. The results show a consistently fast, deadbeat response for step displacement inputs.

No difficulties are anticipated with second- and third-order systems when the principles are applied to physical systems, or only first and second derivative signals are required. However, some difficulties are encountered with higher order systems due to the noise levels in the second, third, and higher derivatives. These difficulties may be avoided for step displacement inputs by not using the higher derivative signals. Feedback loops utilizing only lower derivatives usually compensate the system so that its step response is overdamped. A switching computer may also be designed without using higher derivatives. For step displacement inputs, a switching hyperplane giving precisely correct switching for any order system may be instrumented using only E and $\dot{E}$ signals. The theory has been verified by computer simulation and by design and test of an actual system. It has been shown that suitable engineering approximations are available for the generation of derivative signals and for the formulation of a practical switching computer.

It should be noted that this type of system is quite insensitive to load disturbances. Since the compensation used is solely of the derivative feedback type no attenuation is introduced when the feedback circuits are closed. Thus, under static conditions, any load disturbance is not only opposed by a maximum forward gain but also by maximum damping. Thus, oscillations are not likely to occur unless the load disturbance is severe enough to make the relay open the damping circuits.

References

1. Transient Analysis of Nonlinearized Single Lag Servomechanisms, J. S. Blumenthal, F. J. Beck. *Proceedings*, First U. S. National Congress of Applied Mechanics, New York, N. Y., June 1951, pp. 155–60.

2. The Use of Nonlinear Feedback to Improve the Transient Response of a Servomechanism, J. B. Lewis. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 71, 1952 (Jan. 1953 section), pp. 449–53.

3. Quasi-Optimization of Relay Servos by Use of Discontinuous Damping, W. L. Harris, Jr., C. McDonald, G. J. Thaler. *Ibid.*, vol. 76, Nov. 1957, pp. 292–96.

4. Quasi-Optimization of Relay Servomechanisms by Use of Stored Energy for Braking, C. McDonald, G. J. Thaler. *Ibid.*, vol. 77, 1958 (Jan. 1959 section), pp. 629–34.

5. Investigation of a Nonlinear Control System, I. Flügge-Lotz, C. F. Taylor. *NACATN 3826*, National Advisory Committee for Aeronautics, Washington, D. C., Apr. 1957.

6. Positive-Feedback Phase-Space Trajectories and Application to Servo Systems, Z. H. Meiksin. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 77, 1958 (Jan. 1959 section), pp. 673–79.

7. Matrix Calculus (book), E. Bodewig. Interscience Publishers, Inc., New York, N. Y., 1959, p. 54.

8. An Investigation of the Switching Criteria for Higher Order Contactor Servomechanisms, I. Bogner, L. F. Kazda. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 73, July 1954, pp. 118–26.

9. Increasing the Speed of Response of Certain Automatic Control Systems by Means of Nonlinear Computer Devices, G. M. Ostrovskii. *Avtomatika i Telemekhanika*, Moscow, USSR, vol. 19, no. 3, Mar. 1958, pp. 208–16.

Optimum Synthesis of Multiport Systems Containing Modulators with Periodic Carriers

J. F. EGAN
NONMEMBER AIEE

G. J. MURPHY
ASSOCIATE MEMBER AIEE

THIS PAPER describes the development of a theory of optimum synthesis for a class of linear multiports, which are characterized by the presence of one or more periodically actuated switching elements (with finite dwell times), which multiplex the transmission channels, and by input signals that are generated by stationary random processes. The theory presented is also applicable to the problem of optimum synthesis of linear multiports in which periodic carriers of arbitrary waveshape are amplitude-modulated by the signals being processed; however, the multiplexing problem, in which the carriers are described by

$$u_{ij}(t) = \begin{cases} 1 & nT \leq t - m_{ij} < nT + h_{ij} \\ 0 & \text{elsewhere} \end{cases}$$

of particular practical importance.

The optimum system is defined, for the purposes of this paper, as the system for which the time-average value of the ensemble-average value of the square of the difference between the actual response and the desired response assumes its minimum value. Since the difference between actual response and desired response is commonly known as the system error, the criterion of performance used here is a time-average-ensemble-average square-error criterion, or simply a Time-Ensemble-Average Square-Error criterion. For brevity, it is henceforth referred to as the TEASE criterion.

De Russo[1] has investigated the problem of optimum synthesis of an ideal sampled-data system comprised of a fixed continuous linear element preceded by a pulsed-network prefilter. Franklin[2] has investigated the problem of optimum synthesis of an ideal sampled-data system comprised of a fixed continuous linear element followed by a pulsed-network post-filter. Westcott[3] and Hsieh and Leondes[4] have formulated a technique for solving the mean-square optimization problem for continuous, time-invariant linear multiport systems. All of these problems are encompassed as special cases of the work presented here.

Mathematical Description of the System

Let us consider a linear system with provision for p input signals, $s_1(t)$, $s_2(t)$, ..., $s_p(t)$, and q responses, $r_1(t)$, $r_2(t)$, ..., $r_q(t)$, of such a nature that the transmission channel from the ith input node to the jth output node can be described by the block diagram in Fig. 1. In that diagram, $r_{ij}(t)$ is the component of the jth response which is due to the ith input signal, $u_{ij}(\tau)$, $v_{ij}(\tau)$ and $w_{ij}(\tau)$ are weight-

Paper 61-749, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Ithaca, N. Y., June 18–23, 1961. Manuscript submitted February 14, 1961; made available for printing April 25, 1961.

J. F. Egan and G. J. Murphy are both of Northwestern University, Evanston, Ill.

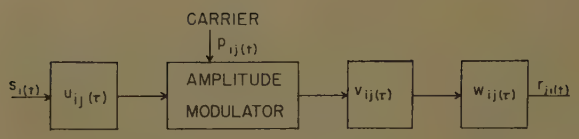


Fig. 1. A block diagram of a single transmission channel of the system under consideration

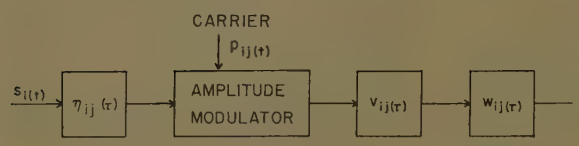


Fig. 2. A block diagram of a system that can be used to generate one term of the left-hand side of equation 19

ing functions of time-invariant linear subsystems, and $p_{ij}(t)$ is a periodic carrier which is amplitude-modulated in the transmission channel from ith input node to jth output node. In this system

$$r_j(t)=\sum_{i=1}^{p}\int_{-\infty}^{\infty}u_{ij}(x)\int_{-\infty}^{\infty}v_{ij}(y)\int_{-\infty}^{\infty}w_{ij}(z)s_i(t-x-y-z)p_{ij}(t-y-z)dzdydx,\ j=1,2,3,\ldots q \quad (1)$$

and

$$R_j(s)=\sum_{i=1}^{p}\{[S_i(s)U_{ij}(s)]*P_{ij}(s)\}V_{ij}(s)W_{ij}(s),\ j=1,2,\ldots q \quad (2)$$

The general optimization problem considered in this paper is most readily described on the basis of a single transmission channel. The ith component of error in the jth response of the system is obtained by subtracting the actual value of the ith component of the jth response from the desired value of the ith component of the jth response. Thus,

$$e_{ji}(t)\triangleq d_{ji}(t)-r_{ji}(t),\ i=1,2,\ldots p;\ j=1,2,\ldots q \quad (3)$$

The error $e_{ji}(t)$ may be due in part to imperfect transmission of the signal and in part to the presence of noise in the signal. To separate these two considerations, let

$$s_i(t)=m_i(t)+n_i(t),\ i=1,2,\ldots p \quad (4)$$

where $m_i(t)$ and $n_i(t)$ denote the message component and the noise component, respectively, of the ith input. (To simplify the work to be presented here, it is assumed that $m_i(t)$ and $n_i(t)$ are stationary random signals with zero mean values.) Also, let $d_{ji}(t)$ be linearly related to $m_i(t)$ through a weighting function $g_{ij}(\tau)$; that is,

$$d_{ji}(t)=\int_{-\infty}^{\infty}g_{ij}(\tau)m_i(t-\tau)d\tau,\ i=1,2,\ldots p;\ j=1,2,\ldots q \quad (5)$$

Then the desired value of the jth response is

$$d_j(t)=\sum_{i=1}^{p}\int_{-\infty}^{\infty}g_{ij}(\tau)m_i(t-\tau)d\tau,\ j=1,2,\ldots q \quad (5a)$$

and the total error in the jth response is

$$e_j(t)=d_j(t)-r_j(t),\ j=1,2,\ldots q \quad (6)$$

Two rather general cases are now to be considered.

CASE 1

The elements which follow the modulator are specified and fixed linear elements that are characterized by time-invariant weighting functions $v_{ij}(\tau)$ and $w_{ij}(\tau)$. Optimum synthesis consists of determining the weighting functions $u_{ij}(\tau)$ of that set of time invariant physically realizable linear subsystems to be located on the input sides of the modulators for which the time average of the ensemble-average squared error (TEASE) E_j, $j=1,2,\ldots,q$, is minimized.

CASE 2

The elements characterized by the time-invariant weighting functions $u_{ij}(\tau)$ and $w_{ij}(\tau)$ are specified and fixed linear subsystems. Optimum synthesis consists of determining the weighting functions $v_{ij}(\tau)$ of that set of time-invariant, physically realizable, linear subsystems to be located immediately following the modulators for which the TEASE E_j, $j=1,2,\ldots,q$, is minimized.

Derivation of an Expression for TEASE

From equations 1, 5, and 6 it follows that

$$e_j^2(t)=\sum_{i=1}^{p}\sum_{k=1}^{p}\left\{\int_{-\infty}^{\infty}u_{ij}(x)dx\int_{-\infty}^{\infty}v_{ij}(y)dy\times\int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}u_{kj}(\beta)d\beta\times\int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda[s_i(t-x-y-z)s_k(t-\beta-\gamma-\lambda)p_{ij}(t-y-z)p_{kj}(t-\gamma-\lambda)]-2\int_{-\infty}^{\infty}u_{ij}(x)dx\int_{-\infty}^{\infty}v_{ij}(y)dy\times\int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta[s_i(t-x-y-z)m_k(t-\beta)p_{ij}(t-y-z)]+\int_{-\infty}^{\infty}g_{ij}(x)dx\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\times m_i(t-x)m_k(t-\beta)\right\}$$

If the required change in the order of several operations can be justified, the time average of the ensemble-average $e_j^2(t)$ can be expressed as

$$E_j=\sum_{i=1}^{p}\sum_{k=1}^{p}\left\{\int_{-\infty}^{\infty}u_{ij}(x)dx\times\int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times\int_{-\infty}^{\infty}u_{kj}(\beta)d\beta\int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\times\int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\psi_{sisk}(x+y+z-\beta-\gamma-\lambda)\phi_{pijpkj}(y+z-\gamma-\lambda)-2\bar{p}_{ij}\int_{-\infty}^{\infty}u_{ij}(x)dx\int_{-\infty}^{\infty}v_{ij}(y)dy\times\int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{simk}(x+y+z-\beta)+\int_{-\infty}^{\infty}g_{ij}(x)dx\times\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{mimk}(x-\beta)\right\}$$

where $\psi_{fg}(\mu)$ denotes the ensemble-average value of $f(t)\,g(t+\mu)$, where

$$\bar{p}_{ij}\triangleq\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}p_{ij}(t)dt=\frac{1}{T}\int_{0}^{T}p_{ij}(t)dt$$

is the average value of the carrier in the channel from the ith input to the jth output, and

$$\phi_{fg}(\tau)\triangleq\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t)g(t+\tau)dt=\frac{1}{T}\int_{0}^{T}f(t)g(t+\tau)dt$$

is the time-average value of $f(t)g(t+\tau)$, i.e., a cross-correlation function of $f(t)$ and $g(t)$.

ptimum Synthesis—Case 1

Let it now be assumed that $v_{ij}(\tau)$ and $_{ij}(\tau)$ are fixed and specified, for $i=1,2,$., p and $j=1, 2, \ldots, q$, and that e optimum physically realizable $u_{ij}(\tau)$ e to be determined. Let it be further sumed that an optimum solution exists, d let this optimum solution be de- ted by $u_{ij}^{\text{opt}}(\tau)$, $i=1, 2, \ldots p$; $j=1, 2,$., q. In general, then, we consider the ss of weighting functions of the form

$$_{ij}(\tau)=u_{ij}^{\text{opt}}(\tau)+\epsilon\eta_{ij}(\tau),\ i=1, 2, \ldots p;\ j=1, 2, \ldots q \quad (11)$$

ere $\eta_{ij}(\tau)$ is an arbitrary nondiscrete, ne-invariant weighting function that is ntinuous at $\tau=0$ and at $\tau=\infty$ and tisfies the conditions

$$_{ij}(\tau)\equiv 0,\ \tau\leq 0 \text{ and } \tau=\infty \quad (12)$$

d ϵ is a real parameter.

Substituting equation 11 into equation gives

$$=\sum_{i=1}^{p}\sum_{k=1}^{p}\Big\{\int_{-\infty}^{\infty}[u_{ij}^{\text{opt}}(x)+\epsilon\eta_{ij}(x)]dx\times \int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times \int_{-\infty}^{\infty}[u_{kj}^{\text{opt}}(\beta)+\epsilon\eta_{kj}(\beta)]d\beta\times \int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\times \psi_{s_is_k}(x+y+z-\beta-\gamma-\lambda)\phi_{p_{ij}p_{kj}} (y+z-\gamma-\lambda)-2\bar{p}_{ij}\int_{-\infty}^{\infty}[u_{ij}(x)+ \epsilon\eta_{ij}(x)]dx\int_{-\infty}^{\infty}v_{ij}(y)dy\times \int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{s_im_k} (x+y+z-\beta)+\int_{-\infty}^{\infty}g_{ij}(x)dx\times \int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{m_im_k}(x-\beta)\Big\} \quad (13)$$

A necessary condition on the optimum ution is that

$$^{j}=0,\ \epsilon=0 \quad (14)$$

Substituting equations 13 and 14 and nplifying yield

$$\sum_{1}\sum_{k=1}^{p}\Big\{\int_{-\infty}^{\infty}\eta_{ij}(x)dx\int_{-\infty}^{\infty}v_{ij}(y)dy\times \int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}u_{kj}^{\text{opt}}(\beta)d\beta\times \int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\psi_{s_is_k} (x+y+z-\beta-\gamma-\lambda)\phi_{p_ip_{kj}} (y+z-\gamma-\lambda)-\bar{p}_{ij}\int_{-\infty}^{\infty}\eta_{ij}(x)dx\times \int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times \int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{s_im_k}(x+y+z-\beta)\Big\}=0 \quad (15)$$

if the order of integration may be changed. Equation 15 can be rewritten as

$$\sum_{i=1}^{p}\int_{-\infty}^{\infty}\eta_{ij}(x)dx\sum_{k=1}^{p}\Big[\int_{-\infty}^{\infty}v_{ij}(y)dy\times \int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}u_{kj}^{\text{opt}}(\beta)d\beta\times \int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\psi_{s_is_k} (x+y+z-\beta-\gamma-\lambda)\phi_{p_{kj}p_{kj}} (y+z-\gamma-\lambda)-\bar{p}_{ij}\int_{-\infty}^{\infty}v_{ij}(y)dy\times \int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{s_im_k} (x+y+z-\beta)\Big]=0 \quad (16)$$

Now, since $\eta_{ij}(\tau)$ satisfies equation 12 and is otherwise arbitrary except for the stated restrictions, the requirement that equation 16 be satisfied is equivalent to the requirement that

$$\sum_{i=1}^{p}\int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times \int_{-\infty}^{\infty}u_{kj}^{\text{opt}}(\beta)d\beta\int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\times \int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\psi_{s_is_k}(x+y+z-\beta- \gamma-\lambda)\phi_{p_{ij}p_{kj}}(y+z-\gamma-\lambda) =\sum_{k=1}^{p}\bar{p}_{ij}\int_{-\infty}^{\infty}v_{ij}(y)dy\times \int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{s_im_k} (x+y+z-\beta),\ x>0;\ i=1, 2, \ldots p \quad (17)$$

A second necessary condition on the optimum solution is that

$$\frac{\partial^2 E_j}{\partial\epsilon^2}>0,\ \epsilon=0 \quad (18)$$

Substitution of equation 13 into 18 yields

$$\sum_{i=1}^{p}\sum_{k=1}^{p}\int_{-\infty}^{\infty}\eta_{ij}(x)dx\int_{-\infty}^{\infty}\eta_{kj}(\beta)d\beta\times \int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times \int_{-\infty}^{\infty}v_{kj}(\gamma)d\gamma\int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\psi_{s_is_k} (x+y+z-\beta-\gamma-\lambda)\phi_{p_{ij}p_{kj}} (y+z-\gamma-\lambda)>0 \quad (19)$$

if the order of integration may be changed. Since each term on the left-hand side of equation 19 may be regarded as the time average of the ensemble average of the square of the output of a system of the form illustrated in Fig. 2, equation 19 is necessarily satisfied, except in the trivial case where all input signals $s_i(t)$, $i=1, 2, \ldots, p$ are identically zero or where the transmission through the system illustrated in Fig. 2 is identically zero for $i=1, 2, \ldots, p$ and $j=1, 2, \ldots, q$. Since equations 17 and 19 constitute a set of necessary and sufficient conditions on the optimum solution, it follows that equation 17 alone is a necessary and sufficient condition except in the trivial cases mentioned.

In general, the set of p equations in equation 17 must be solved simultaneously for the optimum set of weighting functions. However, if $s_i(t)$ and $s_k(t)$ are uncorrelated for $k\neq i$, then equation 17 reduces to

$$\int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times \int_{-\infty}^{\infty}u_{ij}^{\text{opt}}(\beta)d\beta\int_{-\infty}^{\infty}v_{ij}(\gamma)d\gamma\times \int_{-\infty}^{\infty}w_{ij}(\lambda)d\lambda\psi_{s_is_i}(x+y+z- \beta-\gamma-\lambda)\phi_{p_{ij}p_{ij}}(y+z-\gamma-\lambda)- \bar{p}_{ij}\int_{-\infty}^{\infty}v_{ij}(y)dy\int_{-\infty}^{\infty}w_{ij}(z)dz\times \int_{-\infty}^{\infty}g_{ij}(\beta)d\beta\psi_{s_im_i}(x+y+z-\beta)=0 \quad x>0;\ i=1, 2, \ldots, p \quad (20)$$

or, equivalently,

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}e^{sx}F_{ji}(s)ds=0,\ x>0; \quad i=1, 2, \ldots p \quad (21)$$

where $F_{ji}(s)$ is defined as the bilateral transform of the left-hand side of equation 20. If the order of integration may be reversed in this Laplace-transform integral, then

$$F_{ji}(s)=U_{ij}^{\text{opt}}(s)\Psi_{s_is_i}(s)[V_{ij}(s)V_{ij}(-s) W_{ij}(s)W_{ij}(-s)]*\Phi_{p_{ij}p_{ij}}(s)- \bar{p}_{ij}V_{ij}(-s)W_{ij}(-s)G_{ij}(s)\Psi_{s_im_i}(s), \quad 1, 2, \ldots, p \quad (22)$$

Now, if

$$\Psi_{ji}(s)\triangleq\Psi_{s_is_i}(s)\{[V_{ij}(s)V_{ij}(-s)W_{ij}(s)\times W_{ij}(-s)]*\Phi_{p_{ij}p_{ij}}(s)\}\ i=1, 2, \ldots, p \quad (23)$$

satisfies the Paley-Wiener criterion for spectrum factorization, then let

$$\Psi_{ji}(s)=\Psi_{ji}^{+}(s)\Psi_{ji}^{-}(s) \tag{24}$$

where $\Psi_{ji}^{+}(s)$ and $\Psi_{ji}^{-}(s)$ are defined (in the usual manner) so that all the poles and zeros of $\Psi_{ji}(s)$ that lie to the left of the imaginary axis are poles and zeros, respectively, of $\Psi_{ji}^{+}(s)$, and all the poles and zeros of $\Psi_{ji}(s)$ that lie to the right of the imaginary axis are poles and zeros, respectively, of $\Psi_{ji}^{-}(s)$. Then it follows that if a solution to the Wiener-Hopf equation, 20, exists, it is

$$u_{ij}^{\text{opt}}(\tau)=\frac{\bar{p}_{ij}}{2\pi j}\int_{-j\infty}^{j\infty}e^{s\tau}\frac{ds}{\Psi_{ji}^{+}(s)}\times \int_0^{\infty}e^{-st}\left[\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\frac{e^{st}[V_{ij}(-s)W_{ij}(-s)G_{ij}(s)\Psi_{s_im_i}(s)]}{\Psi_{ji}^{-}(s)}ds\right]dt$$
$$i=1,2,\ldots,p,\quad j=1,2,\ldots,q \tag{25}$$

Optimum Synthesis—Case 2

Let it now be assumed that $u_{ij}(\tau)$ and $w_{ij}(\tau)$ are fixed and specified, for $i=1, 2, \ldots, p$ and $j=1, 2, \ldots, q$, and that the optimum physically realizable $v_{ij}(\tau)$ are to be determined. Let it be further assumed that there exists an optimum solution and let this optimum solution be denoted by $v_{ij}^{\text{opt}}(\tau)$, $i=1, 2, \ldots, p$, $j=1, 2, \ldots, q$. It can then be shown (subject to the limitations mentioned in the discussion of Case 1) that

$$\sum_{k=1}^{p}\int_{-\infty}^{\infty}v_{kj}^{\text{opt}}(\gamma)d\gamma\int_{-\infty}^{\infty}u_{kj}(\beta)d\beta\times \int_{-\infty}^{\infty}w_{kj}(\lambda)d\lambda\int_{-\infty}^{\infty}u_{ij}(x)dx\times \int_{-\infty}^{\infty}w_{ij}(z)dz\psi_{s_ks_i}(\beta+\gamma+\lambda-x-y-z)\phi_{p_{kj}p_{ij}}(\gamma+\lambda-y-z)$$
$$=\sum_{k=1}^{p}\bar{p}_{ij}\int_{-\infty}^{\infty}u_{ij}(x)dx\times \int_{-\infty}^{\infty}w_{ij}(z)dz\int_{-\infty}^{\infty}g_{kj}(\beta)d\beta\psi_{s_im_k}(x+y+z-\beta)$$
$$y>0;\ i=1,2,\ldots,p \tag{26}$$

and that, if $s_i(t)$ and $s_k(t)$ are not correlated for $k\neq i$,

$$v_{ij}^{\text{opt}}(\tau)=\frac{\bar{p}_{ij}}{2\pi j}\int_{-j\infty}^{j\infty}e^{s\tau}\frac{1}{\Gamma_{ji}^{+}(s)}\times\left[\int_0^{\infty}e^{-st}\frac{dt}{2\pi j}\int_{-j\infty}^{j\infty}\frac{e^{st}[U_{ij}(-s)W_{ij}(-s)G_{ij}(s)\Psi_{s_im_i}(s)]}{\Gamma_{ji}^{-}(s)}ds\right]ds$$
$$i=1,2,\ldots,p;\quad j=1,2,\ldots,q \tag{27}$$

where

$$\Gamma_{ji}(s)\triangleq\{[\Psi_{s_is_i}(s)U_{ij}(s)U_{ij}(-s)]*\Phi_{p_{ij}p_{ij}}(s)\}W_{ij}(s)W_{ij}(-s)$$
$$i=1,2,\ldots,p \tag{28}$$

and $\Gamma_{ji}^{+}(s)$ and $\Gamma_{ji}^{-}(s)$ are related to $\Gamma_{ji}(s)$ as $\Psi_{ji}^{+}(s)$ and $\Psi_{ji}^{-}(s)$, respectively, are related to $\Psi_{ji}(s)$.

Semifree Configuration—Illustrative Example

The application of the approach, presented in the preceding sections of this paper, to a periodically switched system is illustrated in the following example.

A block diagram of the system under consideration is shown in Fig. 3(A). The switch arm is understood to be in contact with terminal A for $nT\leq t<(2n+1)T/2$ and in contact with terminal B for $(2n+1)T/2\leq t<(n+1)T$, for $n=0, \pm1, \pm2, \ldots$. The optimum weighting function for the system in the absence of the noise $n(t)$ is $g(\tau)$. Both $u_{11}(\tau)$ and $v(\tau)$ are fixed and specified, and the problem is to determine the optimum physically realizable $u_{21}(\tau)$. Since not all the $u_{ij}(\tau)$ are free to be specified by the designer, the configuration is only semifree.

The block diagram is redrawn in Fig. 3(B) to conform with Fig. 1. The carriers employed are

$$p_{11}(t)=\sum_{n=-\infty}^{\infty}\left[u_{-1}(t-nT)-u_{-1}\left(t-\frac{2n+1}{2}T\right)\right] \tag{2}$$

and

$$p_{21}(t)=\sum_{n=-\infty}^{\infty}\left\{u_{-1}\left(t-\frac{2n+1}{2}T\right)-u_{-1}[t-(n+1)T]\right\} \tag{3}$$

where $u_{-1}(t)$ denotes the unit step function with discontinuity at $t=0$.

To find the optimum solution, $u_{21}(\tau)$ is replaced by $u_{21}(\tau)+\epsilon\eta(\tau)$, and the usual procedure is followed. Thus, it is found that a necessary and sufficient condition on the optimum physically realizable $u_{21}^{\text{opt}}(\tau)$ is that it satisfy the equation

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}e^{sx}F(s)ds=0,\quad x>0 \tag{4}$$

where

$$F(s)=U_{11}(s)\Psi_{ss}(s)\{\Phi_{p_{11}p_{21}}(s)*[V(s)V(-s)]\}+U_{21}^{\text{opt}}(s)\Psi_{ss}(s)\times\{\Phi_{p_{21}p_{21}}(s)*[V(s)V(-s)]\}-2\bar{p}_{21}G(s)V(-s)\Psi_{sm}(s) \tag{5}$$

Now, it can be readily shown that

$$\bar{p}_{21}=\frac{1}{2} \tag{6}$$

that

$$\phi_{p_{21}p_{21}}(\tau)=\frac{1}{4}+2\sum_{n=0}^{\infty}\frac{\cos[(2n+1)\Omega\tau]}{(2n+1)^2\pi^2} \tag{7}$$

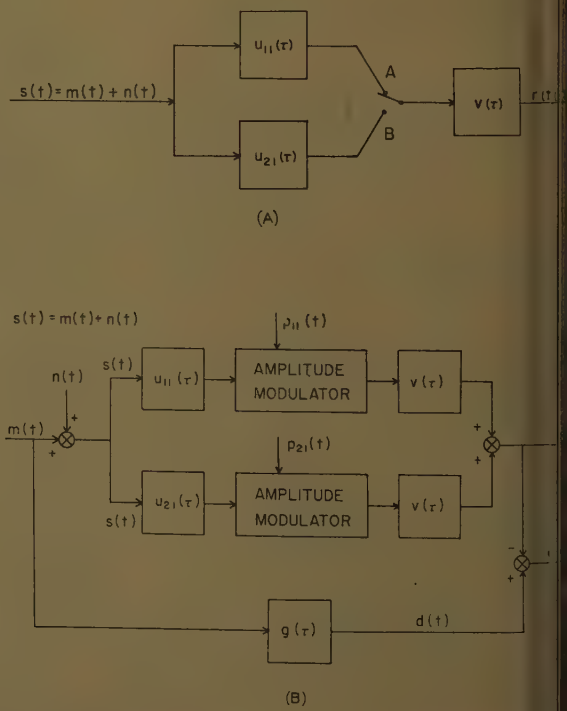


Fig. 3. Block diagrams of the system considered in the illustrative example

here

$$\triangleq \frac{2\pi}{T} \tag{35}$$

the switching frequency, and that

$$_{p_{21}p_{21}}(\tau) = \frac{1}{2} - \phi_{p_{21}p_{21}}(\tau) \tag{36}$$

follows that

$$(s) = U_{21}^{\text{opt}}(s)\Psi_{ss}(s)\{\Phi_{p_{21}p_{21}}(s)* [V(s)V(-s)]\} + U_{11}(s)\Psi_{ss}(s)\times \frac{V(s)V(-s)}{2} - U_{11}(s)\Psi_{ss}(s) \{\Phi_{p_{21}p_{21}}(s)*[V(s)V(-s)]\} - G(s)V(-s)\Psi_{sm}(s) \tag{37}$$

Dividing both sides of equation 37 by $_{ss}^{-}(s)\ \{\Phi_{p_{21}p_{21}}(s)*[V(s)V(-s)]\}^{-}$ gives

$$\frac{F(s)}{_{s}^{-}(s)\{\Phi_{p_{21}p_{21}}(s)*[V(s)V(-s)]\}^{-}} = U_{21}^{\text{opt}}(s)\Psi_{ss}^{+}\{\Phi_{p_{21}p_{21}}(s)* [V(s)V(-s)]\}^{+} + \frac{U_{11}(s)\Psi_{ss}^{+}(s)V(s)V(-s)}{2\{\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)]\}^{-}} - U_{11}(s)\Psi_{ss}^{+}(s)\{\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)]\}^{+} - \frac{G(s)V(-s)\Psi_{sm}(s)}{\Psi_{ss}^{-}(s)\{\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)]\}^{-}} \tag{38}$$

Since, by virtue of equation 31, $F(s)$ s no poles to the left of the imaginary is, the function on the left-hand side of uation 38 has no poles to the left of the aginary axis. If $U_{11}(s)$ has no poles , nor to the right of, the imaginary is, then all the poles of the first and ird terms on the right-hand side of uation 38 are in the left-hand half of e s-plane. Therefore, we require that

$$_{21}^{\text{opt}}(s) = U_{11}(s) + \left\{ G(s)V(-s)\Psi_{sm}(s) - \frac{1}{2}U_{11}(s)\Psi_{ss}(s)\times \frac{V(s)V(-s)}{\Psi_{ss}^{-}(s)\{\Phi_{p_{21}p_{21}}(s)*[V(s)V(-s)]\}^{-}} \right\}_{+} \Big/ \Psi_{ss}^{+}(s)\{\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)]\}^{+} \tag{39}$$

here $\{H(s)\}_{+}$ is used to denote, as usual, e sum of those terms in the partial-action expansion of $H(s)$ that stem from les to the left of the imaginary axis.

It can be shown that

$$_{p_{21}p_{21}}(s) * [V(s)V(-s)] = \frac{V(s)V(-s)}{4} + 2Re\sum_{n=0}^{\infty} \frac{\{V[s-j(2n+1)\Omega]\}\times\{V[-s-j(2n+1)\Omega]\}}{(2n+1)^2\pi^2} \tag{40}$$

here $ReQ(s)$ is used to denote, as usual, e real part of the function $Q(s)$. If $V(s)$ has low-pass characteristics with no appreciable reasonance, the number of significant terms in the series on the right-hand side of equation 40 can be determined largely on the basis of the coefficients $2/[(2n+1)^2\pi^2]$. For $n=0$, 1, and 2, these coefficients are 0.2022, 0.0225, and 0.0009, respectively. Therefore, only the first term ($n=0$) is retained, with the result

$$\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)] \cong \frac{V(s)V(-s)}{4} + \frac{2}{\pi^2} Re[V(s-j\Omega)V(-s-j\Omega)] \tag{41}$$

The use of the solution presented in equation 39 is illustrated in the following example. Let

$$V(s) = \frac{\sqrt{10}}{s+5} \tag{42}$$

$$U_{11}(s) = \frac{1}{s+0.5} \tag{43}$$

$$G(s) = 1 \tag{44}$$

$$\Psi_{mm}(s) = \frac{2}{1-s^2} \tag{45}$$

$$\Psi_{nn}(s) = \frac{20}{100-s^2} \tag{46}$$

$$\Psi_{mn}(s) = 0 \tag{47}$$

and

$$\Omega = 10 \tag{48}$$

Then

$$\Psi_{ss}^{+}(s) = \frac{\sqrt{22}(s+\sqrt{10})}{(s+1)(s+10)} \tag{49}$$

$$\Psi_{ss}^{-}(s) = \frac{\sqrt{22}(-s+\sqrt{10})}{(-s+1)(-s+10)} \tag{50}$$

and, on the basis of equation 41,

$$\{\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)]\}^{+} \cong \frac{0.95\sqrt{5}(s+6.9+j7.5)(s+6.9-j7.5)}{(s+5)(s+5+j10)(s+5-j10)} \tag{51}$$

and

$$\{\Phi_{p_{21}p_{21}}(s) * [V(s)V(-s)]\}^{-} \cong \frac{0.95\sqrt{5}(-s+6.9+j7.5)(-s+6.9-j7.5)}{(-s+5)(-s+5+j10)(-s+5-j10)} \tag{52}$$

Substitution into equation 39 and routine reduction of the resulting expression result finally in the solution

$$U_{21}^{\text{opt}}(s) = \frac{1}{s+0.5} + \frac{0.232(s+5+j10)(s+5-j10)\times(s^3+15.5s^2+55.6s+20.1)}{(s+3.16)(s+0.5)(s+6.9+j7.5)\times(s+6.9-j7.5)} \tag{53}$$

As is sometimes the case in the optimum synthesis of a linear continuous (i.e., non-pulsed) system on the basis of the mean-square error criterion, even though the optimum transfer function obtained possesses no poles on, or to the right of, the imaginary axis, it can be physically realized only approximately, because the degree of its numerator exceeds that of its denominator. However, by the introduction of an additional real pole, at a sufficient distance from the origin, the optimum transfer function can be approximated to any degree desired.

Conclusions

A method for the optimum synthesis of a class of linear multiports containing modulators with periodic carriers has been presented. If the minimization of the Time-Ensemble-Average of the Squared Error (TEASE) is adopted as the goal, the use of this technique leads to the synthesis of the optimum time-invariant, physically realizable, linear subsystems which precede or which follow the modulators. A notable example of the application of this theory is in the optimum synthesis of prefilters for a multiplex data transmission system.

In this paper, a set of simultaneous integral equations is derived as a condition for optimization in the case where correlation between different input functions is not identically zero. These integral equations can be transformed into a set of simultaneous linear algebraic equations by application of the bilateral Laplace transform. To insure physical realizability of the solution, the concept of spectrum factorization is applied to the matrix of coefficients of the set of algebraic equations. A complete description of the solution is quite lengthy; however, the procedure is similar to that followed by Hsieh and Leondes in dealing with continuous systems. The interested reader is therefore referred to their excellent paper.[4]

It is evident that if in the finite-pulsed multiport the area under each rectangular carrier pulse is unity and the pulse width approaches zero, the results presented in this paper are directly applicable to ideal sampled-data multiports. It is of interest to note that, generally, the use of the familiar z-transform does not result in a simplification of the expressions presented and fails to facilitate an extension of De Russo's work to the generalized multiport system.

Possibilities for future work include consideration of the problem of simultaneously optimizing the pre-filters and the post-filters of a finite-pulsed multiport, which has been done for ideal single-channel sampled-data systems,[5] and of

the synthesis of feedback multiports with arbitrary periodic carriers.

References

1. Optimum Linear Filtering of Signals Prior to Sampling, **P. M. De Russo.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 79, 1960 (Jan. 1961 section), pp. 549–55.

2. Linear Filtering of Sampled Data, **G. Franklin.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. IV, vol. 3, 1955, pp. 119–28.

3. Design of Multivariable Optimum Filters, **J. H. Westcott.** *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 80, 1958, pp. 463–67.

4. On the Optimum Synthesis of Multip Control Systems in the Wiener Sense, H. **Hsieh, C. T. Leondes.** *Convention Record*, Insti of Radio Engineers, pt. IV, vol. 7, 1959, pp. 18–

5. Optimum Transmission of Continu Signal over a Sampled Data Link, S. S. **Chang.** *AIEE Transactions*, pt. II (*Applicat and Industry*), vol. 79, 1960 (Jan. 1961 secti pp. 538–42.

Locomotive Repair Costs and Their Economic Meaning to the Railways of the United States

H. F. BROWN
FELLOW AIEE

LOCOMOTIVE REPAIR costs are the largest single item of road or line-haul locomotive operating expense. They are discussed without reference to other operating costs, except depreciation, which should be closely related.

The relationship of repair costs to the age of equipment for steam, electric, and diesel locomotives is shown. Diesel repair costs are higher and rise more rapidly with age than those of the other two types.

Repair costs increase and depreciation charges decrease with service life. The age-point where the sum of these two costs is at a minimum is the economic life of the equipment. Service life can be somewhat longer than economic life without too much loss in operating economy.

The relatively short economic life of the diesel locomotive, despite its higher initial cost, is emphasized and compared with the lower initial cost and longer life of both steam and electric locomotives.

The greater rate of rise in repair costs and the shorter economic life of the diesel locomotive indicate that the Class I railways of the United States will face a serious financial problem within the next 6 years. Almost 80% of their present motive power will be due for replacement by the end of this period at present rate of utilization.

The necessity for a more economic type of motive power is a challenge to locomotive manufacturers and a problem for serious study on the part of the railways.

The Importance of Repair Costs

Locomotive repair costs are of particular importance in the study of railway economics. In the United States for the past 35 years, these costs have been the largest item of operating expense for motive power in road service on Class I railways. This has been true regardless of the changes in types of motive power. Only in 1943 and from 1947 to 1950 were repair costs relegated to second place by fuel costs. The *Statistics of Railways of the United States*, published annually by the ICC (Interstate Commerce Commission), supports these statements.

The items of road locomotive operating expense in the ICC statistics are given in their relative order of importance:

1. Repair costs.
2. Fuel costs, including electric power where used.
3. Wages of engine crew.
4. Depreciation.
5. Enginehouse expenses.
6. Lubricants.
7. Other locomotive supplies.
8. Water.

These constitute all items of locomotive operating expense as defined by the ICC.

This paper discusses repair costs for road locomotives only, and the intimate relationship that should exist between these costs and depreciation charges.

Road locomotive repair costs from 1920 to 1957 are shown in Table I, together with total railway operating expenses. Total repair costs are also shown as proportionate parts of total railway operating expense, and of total railway operating revenue. The relative magnitude of the eight items of locomotive operating expense is shown in Fig. 1 for the year 1957, which was the latest year of published ICC statistics when this paper was written.

Locomotive repair costs measurable in dollars are not comparable from year to year because of the general reduction in the purchasing power of the dollar. Even with a stable dollar value, these costs must be related to the work performed by, and the capacity and age of the locomotive.

By converting dollar costs into a proportionate part of the total railway operating expense, as in Table I, the inflation factor for each year appears almost equally in both the numerator and denominator of this ratio, and, within the accuracy of the figures used, is cancelled out, thus allowing comparison from year to year.

A discussion of repair costs must necessarily start with steam locomotives. Although these may be of but academic interest to the railways of the United States today, nevertheless, steam locomotives have been in operation, in gradually decreasing numbers since 1924,

Table I Repair Costs for Road Locomotives on Class I Railways—In Dollars, from ICC Statistics, and as Proportionate Part of Total Railway Operating Expense (TROE) and Total Railway Operating Revenue (TROR)

	Costs in Millions of Dollars							
	Repairs						Total Repair Costs as Proportion of	
Year	Steam	Diesel	Other	Total	TROE	Operating Ratio	TROE	TROR
1920	512.0		4.1	516.1	5,831	94.36	0.0886	0.0835
1925	388.5		2.7	391.2	4,540	74.10	0.0864	0.0640
1930	293.5		3.8	297.3	3,931	74.43	0.0757	0.0563
1935	191.6		4.0	195.6	2,593	75.11	0.0755	0.0567
1940	230.1		8.0	238.1	3,089	71.90	0.0770	0.0553
1941	284.4		10.5	294.9	3,664	68.53	0.0805	0.0551
1942	342.5		14.0	356.5	4,601	61.63	0.0775	0.0477
1943	411.7		19.0	430.7	5,657	62.48	0.0762	0.0476
1944	474.2		28.3	502.5	6,282	66.57	0.0801	0.0533
1945	461.8		34.4	496.2	6,418	72.10	0.0774	0.0558
1946	451.7		40.9	492.6	6,357	83.35	0.0775	0.0645
1947	453.5		59.9	513.4	6,797	78.27	0.0755	0.0591
1948	441.4		89.7	531.1	7,472	77.26	0.0710	0.0548
1949	334.9		126.0	460.9	6,892	80.32	0.0669	0.0537
1950	316.4		162.7	479.1	7,059	74.52	0.0680	0.0506
1951	326.9		215.6	542.5	8,041	77.39	0.0676	0.0523
1952	243.9		265.5	509.4	8,053	76.11	0.0633	0.0482
1953	173.5		303.5	477.0	8,135	76.29	0.0587	0.0447
1954	86.4	299.6	16.6	402.6	7,384	78.80	0.0546	0.0430
1955	65.3	323.3	18.2	406.8	7,646	75.66	0.0532	0.0403
1956	55.0	365.0	19.8	439.8	8,108	76.85	0.0542	0.0417
1957	30.9	377.4	20.7	429.0	8,228	78.42	0.0522	0.0410

Paper 60-599, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the ASME-AIEE Railroad Conference, Pittsburgh, Pa., April 20–21, 1960. Manuscript submitted January 5, 1960; made available for printing March 14, 1961.

H. F. Brown is with Gibbs & Hill, Inc., New York, N. Y.

The author gratefully acknowledges the editorial assistance of A. G. Oehler and J. Stair, Jr., in preparing this paper.

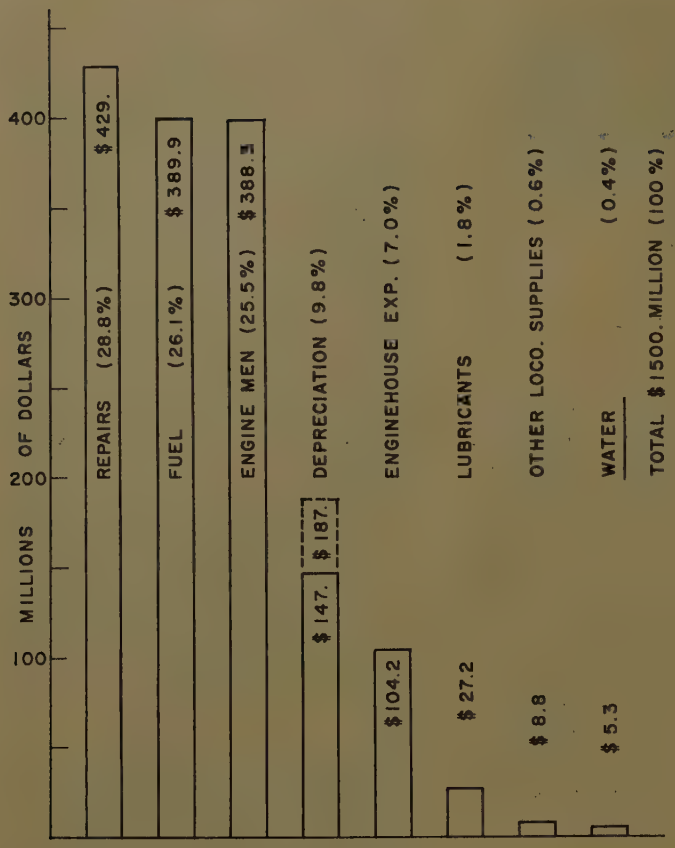


Fig. 1. Relative magnitude of operating expenses for road locomotives, all Class I railways, for 1957

on the Class I railways all during the period under review, and are still operating in limited numbers.

Locomotive Repair Costs

Steam

Steam locomotives rapidly increased in capacity from 1915 to 1940. To determine the economics of these larger locomotives compared with the smaller units built prior to 1915, the unit of costs per locomotive-mile became meaningless.

Considerable research on steam locomotive repair costs was done between 1927 and 1932 by Thomas R. Cook,[1–4] who was with the Baldwin Locomotive Works. Cook worked with the unit of "cost per hp (horsepower) mile" and used this unit in terms of 10,000 hp-miles in order to give the unit costs in dollars. This hp-mile unit, in terms of 1,000 hp delivered to the rail or rim of the driving wheels, has been used in this paper because horsepower rating varies with the different types of locomotives. The 1,000 rail hp-mile unit gives cost in cents, which is often more readily combined or compared with other costs.

Obviously, many other units for measuring repair costs can be used, such as cost per unit of fuel used or cost per ton-mile hauled, but these are not so readily obtainable from the available statistics.

Cook was also one of the first to show the necessity of relating repair costs to the age and capacity of the locomotive. Repair costs increase as the locomotive ages for the same service performed. This important fact is confirmed by other authorities on steam locomotive costs.[5]

In June 1934, the Federal Coordinator of Transportation published in his report the diagram shown in Fig. 2.[6] These costs were obtained from a 1927–1929 survey made of approximately 66% of all types and sizes of steam locomotives then in service. This diagram has sufficient authenticity to be considered as a datum to which repair costs of other types of locomotives may be compared.

Electric

From 1906 to 1944, electric locomotives were applied in increasing numbers on the railroads of the United States. A maximum of 711 were in road service by 1944. The majority had gear drive between the motor and the driving axle. Some were gearless, with motor armatures directly on the axles, or on quills; others were side-rod driven, similar to steam locomotives. There was great variation in the service performed, from short runs in terminals and in tunnels hauling non-working steam locomotives with their trains, to extremely heavy hauling on mountain grades. Consequently, repair costs varied widely.

Several railways used electric locomotives for general operating economy. This service was similar, in every way, to that of the average steam locomotive.

The repair costs considered in this paper are those of two carriers operating gear-driven electric locomotives of similar types, built between 1931 and 1943. They were used for freight and passenger service over rolling profiles, and replaced steam locomotives formerly used in this service.

The author first became interested in comparative electric and diesel-electric locomotive repair costs during the period 1945–1950. In 1953 he collaborated on a comparative study of these costs related to one of these carriers, involving nearly 100 electric and 300 diesel-electric locomotives. The trend of these costs for electric locomotives built since 1931, in terms of 1953 costs, is indicated by *A* in Fig. 3 (and for diesels by *A* in Fig. 6).

In 1955, an opportunity was offered to study electric locomotive repair costs, as related to age, on a second carrier. The trend of these costs is shown by *B* in Fig. 3. All of these electric locomotives had large-diameter driving wheels, with axle arrangements similar to those of steam locomotives, having one or more idle guiding axles.

After diesel locomotives proved that high speeds could be safely attained with smaller driving wheels, without idle guiding axles, and higher tractive forces with all weight on drivers, a few electric locomotives were built incorporating these features. T. F. Perkinson presented the repair costs of such locomotives in a recent AIEE paper.[7] These costs, which are converted into the units and 1953 costs used in this discussion, are shown by the dots and *C* of Fig. 3. These locomotives are used for extremely heavy freight haulage on mountain grades.

Diesel-Electric

Less than 100 diesel-electric locomotives were in road service on Class I railways of the United States prior to 1940. After 1945, however, diesels were acquired in large numbers to replace the worn-out steam locomotives, 40% of which were of pre-1915 vintage. Relatively few steam locomotives had been built since 1930 because of the depression and World War II.

About 20% of the diesels now in service were acquired prior to 1949 and more than 60% were acquired between 1949 and 1953. Today there are nearly 20,000 diesel-electric units in road service. In 1957, approximately 2.41 diesel units made up the average road locomotive.

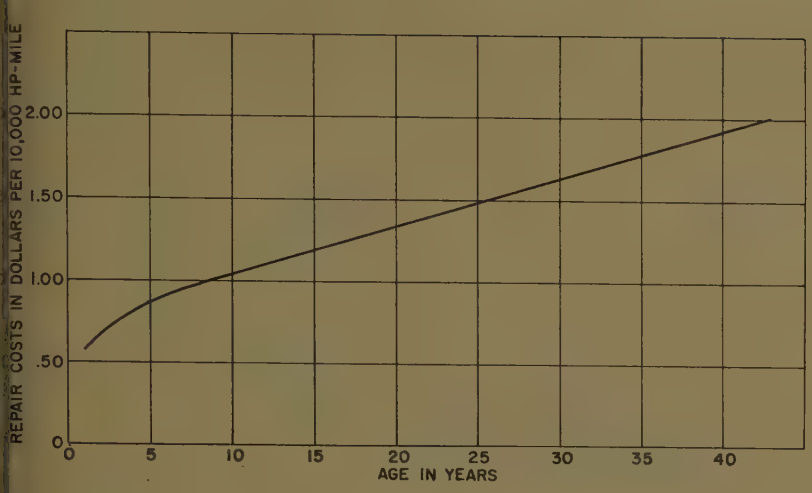


Fig. 2. Cost of steam locomotive repairs in dollars per 10,000 hp-mile unit. From Report of Federal Co-ordinator of Transportation, June 1934. Costs are approximately 1929 price level (Fig. from reference 6)

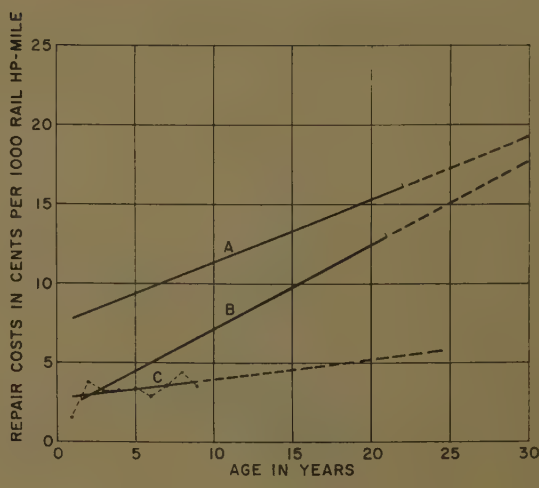


Fig. 3. Electric locomotive repair costs on railways A, B, and C, 1953 price level

It was apparent early in the use of diesels locomotives that repair costs were not following the pattern set by steam and electric motive power, but were more cyclic, due to the heavy repairs which were periodically required by the internal-combustion-type prime mover. To smooth out these periodic high points, manufacturers suggested that repair costs be kept on a cumulative basis. These were compiled by adding the repair costs for each period, either month or year, to the accumulated total for all preceding periods, and dividing the total by the total mileage performed, or by the number of elapsed periods. This method was useful as it related the repair costs to the age of specific groups of locomotives; it is also useful in the determination of the economic life, as will be shown. Cumulative costs are misleading, however, because they show only the average or one-half of the actual rise in repair costs incurred,[8] and cannot be directly compared with similar costs of other motive power, not kept in this manner, unless the rise in cumulative costs is multiplied by two. It is possible that many railways have accepted these cumulative costs for the actual costs incurred.

One manufacturer made elaborate and painstaking efforts to compile diesel-electric locomotive repair costs on a number of carriers during the period 1949–1954. However, additional diesel units were being continually acquired in such large numbers by the railways studied that the "average age" remained nearly constant between 2 and 4 years for this period, and the cost was more or less constant between 20 and 22 cents per unit-mile. Such statistics were consequently of little value in the determination of rise in repair costs with age.

Committee 16 of the American Railway Engineering Association made similar studies in 1949.[9] Their report indicated that repair costs did rise with the age, but made no determination of definite trends in these costs. The age of the road diesel locomotives studied was not greater than 69 months.

In 1955 the author collaborated on an engineering study undertaken for a large Class I carrier to determine the economic life of diesel-electric locomotives, based on a study of repair costs, for the purpose of establishing realistic depreciation rates for tax purposes. The diesels operated by this carrier, with minor exceptions, were less than 5 years old. Accurate repair costs related to age were obtained from several other large carriers which had been operating diesels for as long as 12 and 14 years. Costs for the 3 consecutive years of 1951 to 1953, were obtained for diesels from 1 to 11 years of age from one of these carriers. This was a large trunk line with appreciable traffic, which operated over generally nonmountainous routes. Fig. 4 shows these 3-year costs related to age, and converted into 1953 costs, in terms of cents per 1,000 rail hp-miles.

Fig. 5 presents similar costs for a 5-year period, in terms of cents per 1,000 rail hp-miles, of another large carrier located in another part of the country. This carrier also had reasonably heavy traffic and operated over moderate profiles.

Additional repair cost information which could be related to the age of the equipment was gathered from several other carriers, and was found to be generally consistent with the results of the major analyses.

The repair costs taken from a paper presented before the Pan American Railway Congress by T. T. Blickle in 1953[10] were also analyzed. They generally conformed to the other analyses. In that paper, however, the costs of repairs, fuel, and lubricants were given as percentages of the total of these costs, which complicated the exact determination of repair costs in dollars. The repair costs in relation to age of more than 3,000 diesel units of all ages were studied, up to and including rebuilding costs. The important trends are graphically presented in Fig. 6. Rebuilding costs were as high as 75 to 90% of the original cost, depending on age and mileage performed. With such costs, rebuilt units, by ICC ruling, must appear on the books as new units and the original units must be retired and written off.

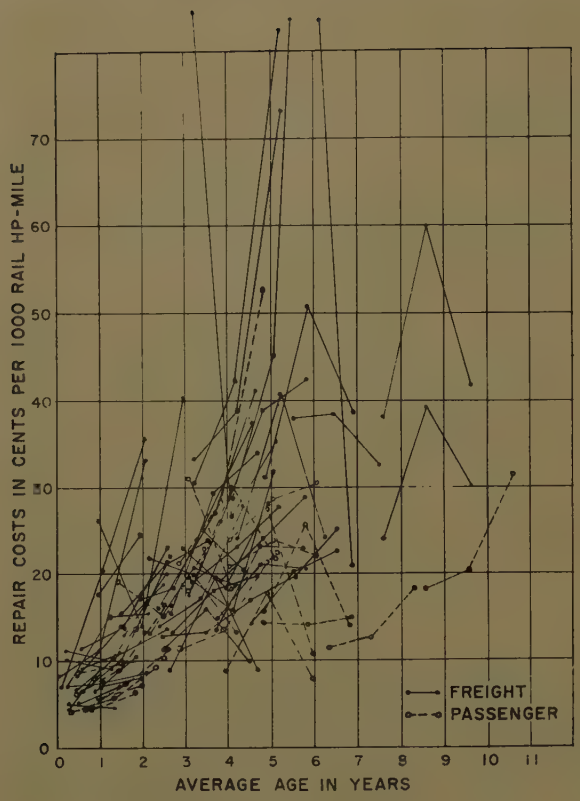
As a result of this study, it was found that repair costs for diesel-electric locomotives did rise with age, but at a much faster rate than for steam or electric locomotives. This pointed to an economic life for road locomotives, in some cases, as low as 8 years, but rarely exceeding 14 years. The average economic life for road diesels appeared to be $12\frac{1}{2}$ years and for diesels in yard service somewhat less than 18 years.

To determine the economic life, depreciation and its relation to repair costs must be considered.

Depreciation

Depreciation, in railway accounting, is an operating charge for the cost of equipment spread over the service life. It should equal, during the life, the original cost reduced by the ultimate scrap value.

By ICC ruling, depreciation of motive power is an item of operating expense under maintenance of equipment. A



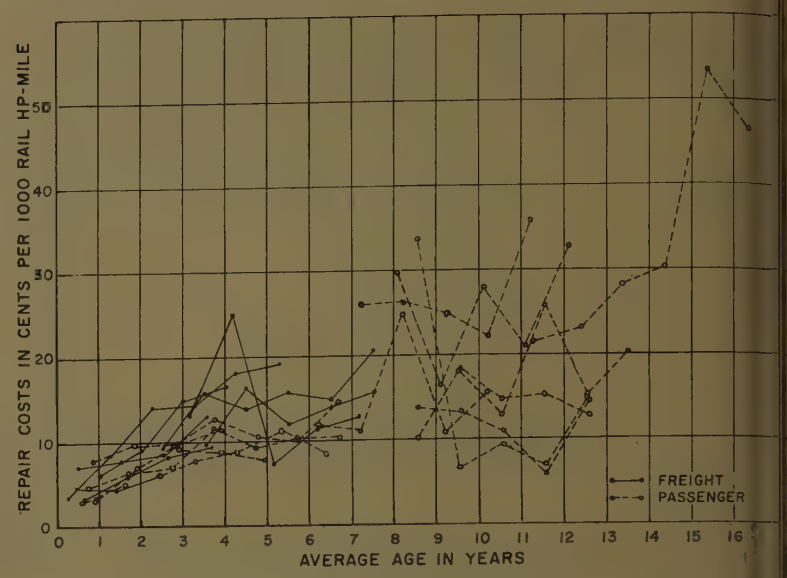


Fig. 4. (left). Repair costs of 742 road diesel locomotives on K railway for three consecutive years, 1953 price level

Fig. 5. (above). Repair costs of 225 road diesel locomotives on L railway for five consecutive years, 1953 price level

realistic depreciation rate is essential to prevent depletion of assets when renewals become necessary. Neither the ICC nor the Internal Revenue Service establishes depreciation rates. Either authority may approve any rates established by the railways, if based on proper supporting data.

Depreciation rates based on a 30-year service life for steam and electric locomotives have been used by the railways for many years. Early in the use of diesel locomotives, depreciation rates—based on a 20-year and a 25-year life, respectively, for road and yard units—were approved by the ICC, keeping in mind the shorter life of the internal-combustion-type engine.

Now, from repair cost studies and from the actual scrapping and rebuilding being done by the various carriers who purchased diesel units before 1947, a more realistic service life seems to be 15 and 20 years, respectively. On this basis depreciation charges of $147 million, shown in Fig. 1 for 1957, would be approximately $187 million.

If a railway's earnings are satisfactory, it is advantageous to have ample depreciation rates, as these charges are a proper deduction, with other operating expenses, before taxes. With small earnings, the tendency is to keep depreciation rates at a minimum, so that, in spite of taxes, the earnings may appear as favorable as possible.

Where depreciation rates are too low, retirements and renewals which are not fully covered by the depreciation reserve must be charged to the Profit and Loss account, or be otherwise reflected by changes in assets on the General Balance Sheet.

Depreciation rates, on a straight-line or average basis for any fixed number of years, may be graphically represented as in Fig. 7. If there is no scrap value, and the service life of a unit of equipment is 20 years, curve A shows that an average of 5% of the cost must be charged to the depreciation account each year during the 20-year period. If the service life is 10 years, then an average of 10% must be charged off each year. Similarly, if the service life is only 5 years, then an average of 20% must be charged off each year. If there is a scrap value, the annual depreciation charges are reduced, and a family of hyperbolic curves, one for each scrap value assumed, can be made, as shown by B, C, and D in Fig. 7.

Since average annual depreciation charges fall and repair costs rise as the service life increases, then, at some point in the service life, the proper combination of these two expenses is at a minimum. That point is the economic life of the equipment.

Since depreciation per year is most frequently expressed as a percentage of the original cost, one method of determining economic life is to convert the known repair costs into per-cent values of the original cost, and to combine graphi

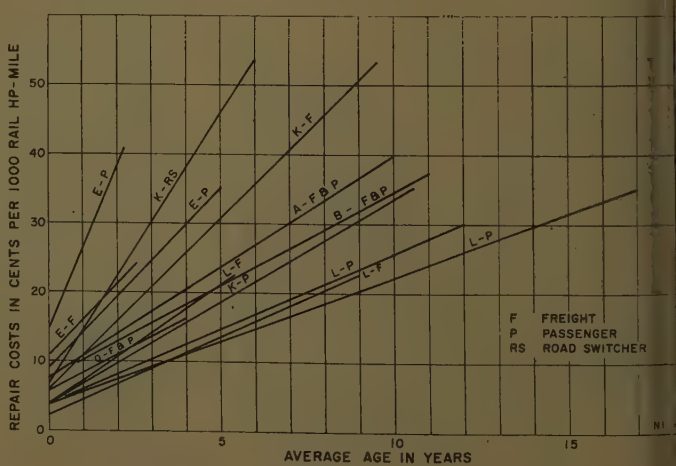


Fig. 6. Repair cost trends for approximately 2,500 road diesel locomotives on railways A, B, E, K, L, and O; 1953 price level

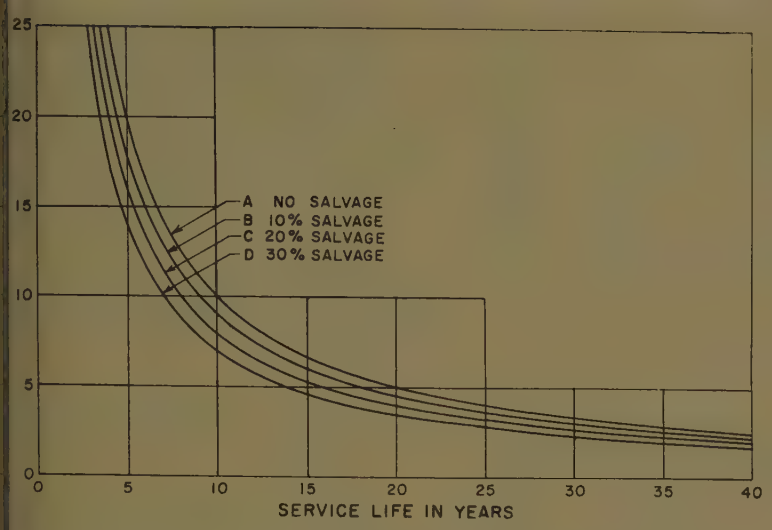
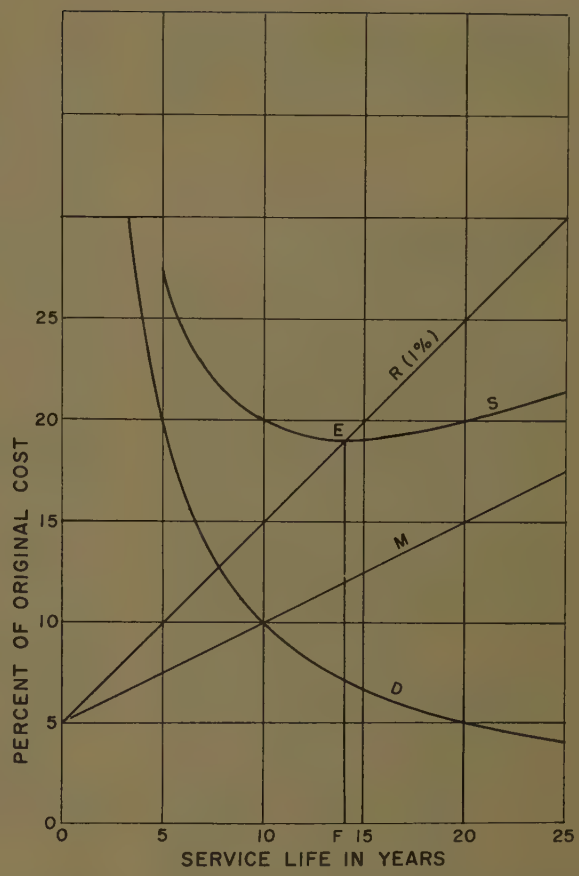


Fig. 7 (above). Average annual depreciation rates for any service life



g. 8 (right). Summation of repair costs (R) and deprecitaion (D). Repair sts having rate of rise of 1% (R) have an average rise (M) of one-half the slope R. The sum of the ordinates of D and M produces the curve S, the low point of which (E) occurs at the economic life

ally the average rate of rise of repair sts with the depreciation curve. The sulting combination curve is a flattened -shape as in Fig. 8 and asymmetrical out the low point, which defines the onomic life. This curve rises more owly after the low point. For this ason, the service life may be as much 20% longer than the economic life, beuse this rise for several more years may be relatively small. However, each point on the U-shaped curve represents the sum of depreciation charges and repair costs, averaged over each year of the service life. Any rise after the low point or economic life is, therefore, cumulative and retroactive in its indication of these costs for the entire service life.

The author treated this subject more fully in a previous paper[11] in which he showed that a straight line, drawn through the point of origin in the diagram of the various hyperbolas for depreciation rates, will intersect any hyperbola at the economic life, provided the straight line has a slope equal to 1/2 the rate of rise of re-

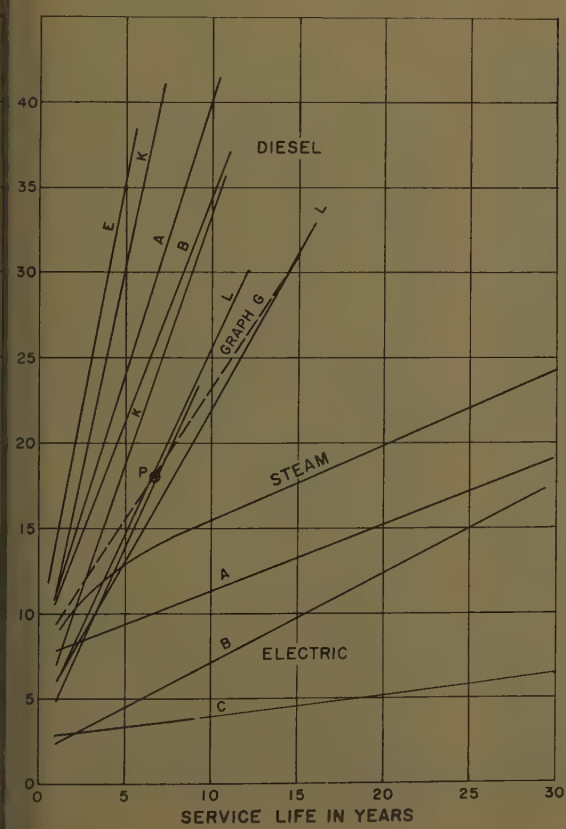
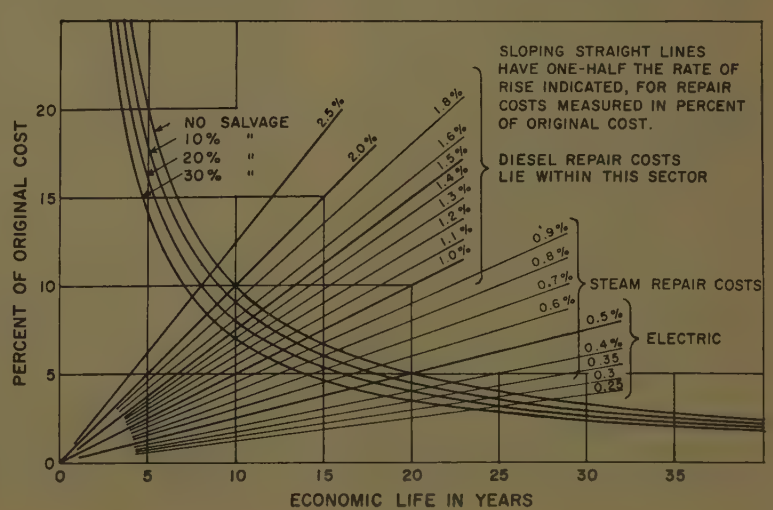


Fig. 9 (left). Repair costs of various types of road locomotives compared, 1953 price levels

Fig. 10 (below). Economic life is determined by the intersection of sloping straight lines (repairs) and hyperbolas (depreciation)



pair costs, measured in percentage of original cost; see Fig. 10. It is therefore apparent that the greater the rate of rise in repair costs, the shorter the economic life.

Comparisons of Motive Power Repair Costs

In Fig. 9, the data from Figs. 2, 3, and 6 for the repair costs of various types of locomotives are shown in terms of 1953 costs. While steam and electric locomotive repair costs have about the same rate of rise, electric costs are much lower than steam. Diesel repair costs are not only higher than the others, but the rate of rise is also much greater. On this diagram, the line *graph G* was calculated for diesel repair costs which would give a service life of 15 years. The actual diesel repair costs for 1957 ($377.4 million) for the Class I railways were converted into cost per 1,000 rail hp-mile, using 1,500 hp as the average road diesel *engine* rating, 82% of which was delivered to the rail. The total mileage performed in road service and the average number of road diesel units were taken from ICC statistics. This result was converted from 1957 dollar costs to 1953 costs to agree with the other cost data shown. The final result was plotted at 6.6 years, the average age of all road diesel units in 1957, at point *P*, which lies almost exactly on *graph G* at 17.8 cents.

All these unit repair costs may be converted into comparative percentage values of the original cost of the locomotives by using the approximate costs that prevailed during the latter part of the period under review:

Steam locomotives: $45 per maximum continuous hp

Electric locomotives: $110 per maximum continuous rail hp

Diesel locomotives: $122 per rail hp or $100 per rated engine hp

The rate of rise in repair costs for each type will depend on the annual mileage, and will fall within the sectors designated in Fig. 10. The economic life of each type is shown by the intersection of the sloping straight lines with the hyperbolas.

Despite the fact that repair costs related to the original equipment costs show a lower rate of rise, and therefore a longer economic life, as the original cost increases, the repair costs of diesel power still show the greatest rate of rise and hence the shortest life.

Theoretically, the economic life should be based on the replacement cost and not on the original cost. It is well known that all costs have been rising continuously during the past 40 years. However, the dollar value of repair costs also has been rising at about the same rate, so that it makes very little difference in this discussion whether original or replacement costs are used. The information shown in Fig. 10 is correct for either cost, within the relative values previously indicated.

Service Life of Motive Power

The information in Fig. 10 indicates that steam locomotives have an average economic life of approximately 21 years. This means a service life of about 25 years without undue drain on operating revenue during the final years of service life. As the number of steam locomotives declined, the average annual mileage became lower, and the service life was extended to 30 years to agree with the average life of the boiler.

Electric locomotives have an economic life of 23 to 25 years, which indicates that the service life is somewhat longer than that of the steam locomotive, as has been proved by experience. Some electric locomotives in the United States have been in service since 1906, and are still serviceable. No comment is offered as to the operating economies involved. They cannot be too unfavorable, however, in view of the service being performed.

The long life of the electric locomotives operated in Switzerland has been mentioned frequently in the technical press.

In striking contrast to the steam and electric locomotive is the relatively short economic and service life of the diesel. It will be noted in Fig. 10 that with a higher scrap value, which might be realized in rebuilding, the life is still shorter.

Since the rebuilding cost is, or should be, less than the original cost, the rate of rise in repair costs after rebuilding, in percentage of rebuilt cost, may be greater than before rebuilding. This would indicate an even shorter economic and service life for rebuilt diesels, unless the rate of rise in repair costs is greatly reduced after rebuilding. This has yet to be proved.

The short life span of the diesel is often defended by stating that it performs each year twice the mileage of an equivalent steam locomotive. Assuming that this statement is true, the total mileage for the service life of each would be approximately the same, with the initial cost of the diesel still twice that of the equivalent steam, and with the average repair costs during the life higher, per 1,000 rail hp-mile, for the diesel.

It is apparent from this discussion th an important and serious problem w face the Class I railways within the n 5 or 6 years, as nearly 80% of the re motive power now in service will be d for replacement at the present rate utilization. Some of this power has s ready reached its economic life. Only small percentage has been fully preciated. The financing problem will a weighty one as more than $2 billion involved.

Apparent Savings Made in Repair Costs

Table I indicates that repair cos shown as a proportion of the total r way operating expense, have declin during the period 1925–1957. The to number of locomotives has declined ab principally because larger steam lo motives came into use following Wo War I and branch-line and short-ha traffic was lost to the highways af 1920. During the business depress from 1930 to 1940 few new locomoti were installed. Consequently, the aver age of all motive power increased dur this period from slightly more than years in 1925 to 27 years in 1945. T reduction in numbers, offset by this crease in average age, kept repair costs a high level during this period.

After World War II, the loss of branc line and short-haul traffic became ac celerated, causing further reduction i train service and the number of lo motives required. The extensive acqui tion of new motive power which has curred since 1945 has reduced the avera age, so that in 1957 this was just bel 10 years. These two, the decline in nu ber and in average age, account most of the reduction in repair co shown since 1945. The change in ty of motive power since 1940 has h little favorable effect on this reducti as a study of Fig. 9 will show.

The railways of the United States pa a substantial penalty in high repair co for allowing their steam motive pow to become so old during the period fr 1925 to 1945.

Conclusions

The diesel engine has been increas materially in rated output during the p 15 years by supercharging, greater f injection, higher mean effective pressur and slightly higher speeds, without ba change in design or in principal par Some parts and materials are now be worked harder than previously. Rep

sts have continued to rise with age at e rates indicated in Fig. 9.

In comparison with other types of tive power, and from the study of re- ir costs, it seems apparent that the nerican-built diesel locomotive, in road vice generally, is:

Too small in horsepower capacity per it

Too heavy in weight per unit for the pacity.

Too expensive in first cost for the rsepower capacity.

Much too expensive to maintain.

Too short in its economic and service e.

Therefore, too great a consumer of ilway capital.

These factors negate the claim that this pe of locomotive is an economic boon American railways.

Multiple units are actually a necessity order that diesels may equal the per- rmance of steam or electric locomotives, hich they have replaced. Although nerally hailed as a virtue, multiple-unit eration greatly multiplies the invest- ent in motive power required per train.

All locomotives must have greater rsepower per ton of train weight for pid acceleration and high speeds than r starting and slower speeds. The diesel unit excels only in the latter requirement.

The best American diesel will deliver no more than 16 hp per ton of locomotive weight, although European manufacturers have developed types capable of producing 30 hp per ton of weight. American manufacturers can build from 150 to 200 hp per ton of weight into an automobile or an airplane, but they are still unable to install in a single diesel locomotive unit the maximum horsepower built into steam locomotives in 1939.

Can the railways afford to perpetuate the consumption of capital required by presently designed diesel locomotives especially where numerous units are necessitated because of dense traffic?

Diesel manufacturers have the next 5 years to develop a diesel that will equal, economically, other types of motive power. Similarly, American railway officials, concerned with traffic density, have the same 5 years to study the real economics of the types of motive power available, both here and abroad.

Types having greater horsepower and faster speeds may permit railways not only to retain the traffic now carried, but possibly to recapture some of that which has been lost to other transportation agencies. Shorter trains and more frequent service might also aid in accomplishing this. The economic future of American railways depends materially on their ability to secure and apply the most economical and longest-lived motive power.

References

1. How New Locomotives Pay for Themselves Through Savings in Maintenance, **Thomas R. Cook.** *Baldwin Locomotives*, Philadelphia, Pa., Apr. 1932.

2. Determination of Savings with Modern Power, **Thomas R. Cook.** *Ibid.*, Oct.

3. The Effect of Operating and Maintenance Savings with Modern Power on Income Account and Cash Position, **Thomas R. Cook.** *Ibid.*, Jan. 1933.

4. The Economic Life of Locomotives and Its Relation to Locomotive Performance and Operating Expense, **Thomas R. Cook.** *Ibid.*, Jan. 1934.

5. Factors Relating to Selection of Type of Locomotive for Various Operations (Steam, Diesel-electric, Electric), **K. Cartwright.** Pan American Railway Congress, June 1953.

6. The Steam Locomotive (book), **Ralph P. Johnson.** Simmons-Boardman Publishing Corporation, New York, N. Y., 1945.

7. Virginian Railway Motor-Generator Electric Locomotive Maintenance Costs, **T. F. Perkinson.** *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 79, Mar. 1960, pp. 33–35.

8. Cumulative Repair Costs, Where Their Fallacy Lies. *Railway Age*, New York, N. Y., Mar. 21, 1955.

9. *Proceedings*, American Railway Engineering Association, vol. 51, 1950, pp. 98–111.

10. The History, Development, Operation and Performance of Diesel Locomotives of the Santa Fe Railway, **T. T. Blickle.** Pan American Railway Congress, June 1953.

11. What's the Life of a Diesel? **H. F. Brown.** *Railway Age*, July 30, 1957.

Discussion

. C. Cross (Westinghouse Electric International Company, New York, N. Y.): t is worthy of note that the data presented y Mr. Brown concerning the rate of rise n maintenance cost and the economic life f road diesel-electric locomotives have ot been controverted by any individual or rganization up to the present time. On he contrary, the relatively short economic fe appears to be accepted fact by both ailroads and locomotive manufacturers. ne manufacturing official has publicly ated that his company believes the verage life of a diesel-electric locomotive n road service is a period of approximately 5 years.

Fig. 9 of the paper shows a comparison f repair costs of various types of road comotives. It is not clear whether the an-wise spread of cost shown for diesel comotives is actually due to variations n locomotive design or to differences in verage annual mileage. It is to be expected that the rate of rise in cost will ncrease in proportion to the annual ileage.

The available data indicate that operated annual mileages, varying from approximately 45,000 miles to 100,000 miles, re accrued on individual locomotives in differing road services.

H. F. Brown: Mr. Cross points out that on exceptions have been taken to my statement regarding the diesel's relatively short economic life. When originally presented at the joint AIEE-ASME meeting in Pittsburgh, Pa., in April, 1960, this paper aroused oral discussion by a representative of one manufacturer in which this short life was questioned.

However, short economic life, demonstrated by the more rapid rate of rise in repair costs, has possibly even better proof in the actual retirements of this type of motive power, compared with acquisitions. The cumulative number of diesel units retired by the class I railways at the end of 1959 was 3,153, according to ICC annual statistics. This was more than the cumulative 3,044 units acquired at the end of 1944, which would indicate an average life of approximately 14.5 years for these retired units.

Since more than 50% of these units were in yard service in 1944, the actual retirements indicate well under 15 years of life for road power and 20 years for yard power. They also confirm the shorter lives indicated by rates of rise in diesel repair costs, shown in Fig. 9. The short life of diesel motive power is no longer a theory—it is a fact that is going to become more impressive within the next five years.

Mr. Cross asks an explanation of the "fan-like spread" of diesel repair costs shown in Fig. 9 (and shown more clearly in Fig. 6.)

A number of factors affect this rate of rise with age, regardless of present design or manufacturer. Although the unit of "per 1,000 rail-horsepower mile" is a somewhat more accurate "yardstick" for measuring repair costs than would be the "per locomotive mile" or "per unit mile," it may not be quite the same for all railroads since there are physical differences in the lines (grades); in traffic requirements (speed); and in shop conditions and practices.

Time factors also are involved with all these costs such as, for example, periodic inspections and overhauls, when the mileage performed is a secondary consideration. In some cases, a greater annual mileage may show a lower rate of rise in repair costs because of fewer time-factor costs. Moreover, the loading of motive power, i.e., whether average performance is at partial or at full-load capacity, has an important bearing on repair costs and their rate of rise.

Fig. 6 seems to indicate that, in general, diesel motive power on the same railroad has a greater rate of rise in these costs in freight than in passenger service. This would indicate that horsepower capacity of the unit has a bearing on these costs, since usually, for passenger service, fewer units of larger horsepower capacity are used in multiple-unit than is the case for freight service.

Steam and electric motive power have the

same fan-like spread in repair costs as has diesel power—and probably for the same reasons—but their general trends are lower, as shown in Fig. 10.

A simple straight-line trend for repair costs may be determined for any period from a plot of these costs related to age and corrected to the desired price level; by inspection, if the plotted points indicate a fairly uniform pattern of rise; or by the "method of least squares," if the points are scattered. Such straight-line trends often show a steeper rate of rise, or slope, for a shorter, earlier period than for a longer or later period. This factor is indicated in Fig. 6.

As repair costs are (theoretically) zero at zero age, if rates of rise diminish as the age progresses, they may produce a definite curved pattern. Such a curve might be represented by the positive values of the parabolic equation

$$y^2 = Ax$$

where y represents the repair costs; x, the age; and A, a constant to be determined. In such a formula, a straight-line trend would be a tangent (or a secant parallel to such a tangent) to this curve at any year, and would explain, in part, the reduced slope for the greater term of years.

Both steam and electric locomotives have been built with capacity in one unit capable of handling the longest trains. Where multiple-unit operation of several smaller-capacity units is required to replace the performance of such larger units, it seems almost unnecessary to point out that many repair costs become multiplied, because numerous electrical as well as mechanical items having wear are doubled, trebled, or quadrupled through multiple-unit operation. This is another important factor contributing to the steeper rise of diesel motive power repair costs. There can be little doubt that the concept of small-capacity motive power units, even though capable of multiple-unit operation by one crew, has added materially to railway operating expense in items in addition to repair costs.

In spite of the change from steam to diesel and the great reduction in number of units caused by loss of traffic, repair costs continue to be the largest single expense item. In 1959, diesel repair costs for road and yard power combined was $451,000,000 on the class I railways. Total investment in this type of motive power was well over 3.5 billion dollars, and replacements of the same type, even at original prices, would call for an average annual expenditure of $233,000,000, based on a maximum of 15 years' useful life. If purchased, as in the past by annual i stallments, the interest at 4% on the unpa balance will average $70,000,000 annual The sum of these three charges is $754,00 000 annually.

Since this expensive motive power mov neither passengers nor freight on the ra today any faster than they are moved on t highways, railways are rapidly losing tra to the automotive vehicles on these hig ways. Significantly, the greatest part t this diesel railway motive power is suppli by the automotive industry itself.

Are railways blind to the fact that unl they procure cheaper, longer-lived, fast and hence more powerful, motive pow within the next decade, many of them w cease to exist because of continuing tra loss? And is it not rather naïve to exp the automotive industry to design and bu such cheaper, more powerful, and long lived motive power, to compete with t traffic on the highways, which after all their proper and larger field of endeavor?

Time is running out. The railways of t United States cannot much longer ign the significance of such high motive pow costs. Nearly every other country in t world has found and is using more econo ical motive power for their dense traffic.

General Principles of the Experimental Equipment Developed by the French National Railroads for Remote-Controlled Railroad Operations

J. C. BLUMSTEIN
NONMEMBER AIEE

THE CONCEPT of automation in the railroad industry is not a new one. It has been applied for some time in Centralized Traffic Control (CTC), for example, and in the operation of classification yards. However, its application to the remote control of train operations is of more recent date.

In the operation of trains by remote control, the engineman is released from any decision and from any control related to the acceleration or deceleration of his train. The actual train movements are directed from a centrally located office where the required orders are prepared and transmitted to the locomotive receiving equipment through an adequate medium.

When the problem of train operation is analyzed, it is seen that safety requires two fundamental and separate actions:

1. The dispatcher, or the signal tower operator under his direction, must prepare the given train itinerary by correctly setting the track switches and signals, taking into account, of course, the movements of any other trains involved.

2. The locomotive engineer must then operate his train in accordance with the signal indications, and the various orders received.

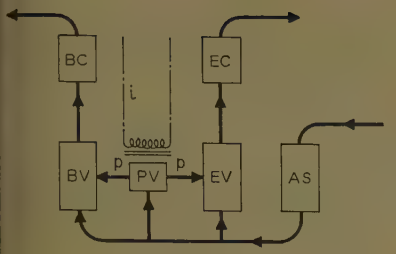
The first part of this procedure has already been automated to a large degree, since interlockings prevent any incorrect moves when a route is being automatica set up from a centrally located tow The second part, however, involving t actual operation of the train, is still co trolled by an engineman in the cab of t locomotive. The information given him by wayside signals is sometimes r peated by signals in the cab, which, unacknowledged or disregarded, can ca the automatic application of the brak Radio has also been used to transmit v bal orders to the engineman while his tra is in motion. Therefore, the basic r quirements for operating a train by r mote control, without the presence of engineman in the locomotive cab, i clude the following:

Fig. 1. Remote-controlled locomotive operation

Paper 61-878, recommended by the AIEE Land Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Railroad Computer and Automation Conference, Cleveland, Ohio, June 6–7, 1961. Manuscript submitted January 24, 1961; made available for printing April 25, 1961.

J. C. Blumstein is with The French National Railroads, New York, N. Y.

The author acknowledges the assistance of M. Laplaiche, Director of Research, Societe Nationale des Chemins de fer Francais, in the preparation of this paper.



2. Electropneumatic servomechanism (simplified diagram)

—Air supply to feed valves at constant pressure of 64 psi
—Pilot feed valve, under control of i
—Engine feed valve, under control of PV, as i exceeds i_0
—Brake feed valve, under control of PV, as i is less than i_0
—Pilot pressure, 0 to 57 psi, regulated by value of i
—Operating current, 0 to 150 milliamperes
—Engine control
—Brake control

The transmission of information and lers from a centrally located point to re- ving devices on the train in motion.

Control of the speed of the locomotive a predetermined rate.

The French National Railroads en- avored to meet these basic require- nts before proceeding further.

The answer to the first of these re- irements was demonstrated on April 18, 55, when an electric train was operated remote control for 10 miles on a main- e track, by orders transmitted from a tant station. To prove their con- ence the engineers locked all the doors the locomotive and allowed no one in e cab during this trial.

A d-c electric locomotive of the *B-B* 00 type, hauling a 5-car train, was rted, accelerated, slowed down, and pped. Orders were transmitted from wayside radio station to the locomotive, d their execution in the correct sequence s observed from a rail car running on cks parallel to those of the experi- ntal train. The results of this test re very promising.

A solution of the second requirement s found by experimenting with a shunt- locomotive, operating in a yard; see g. 1. In this test, the speed of the omotive was entirely controlled from e switch tower over the range from 0 to 25 mph.

This test was limited to a shunting omotive to avoid the difficulties in- lved with the deceleration of long ins having pneumatic brakes. Re- arch is being carried out, however, ward developing electropneumatic brake equipment which will solve this problem. The application to a shunting locomotive was made also because it was expected, and demonstrated by the tests, that when the locomotive is controlled from the switch tower, the efficiency of the operation is increased to a fair degree.

The six aims of the French National Railroads were:

1. To control the shunting locomotive by radio, usually from the general yard tower. In the initial program, control was planned for three different speeds: 15 mph, 9 mph, and the humping speed of 2.4 mph, with slight possible variations around this value. Later, it was thought more desirable to achieve continuous control over the entire range of speed between 0 and the maximum of 25 mph.

2. To direct movements, both forward and reverse, by remote control.

3. To provide means for an adequate approaching speed so that the locomotive would escape the tower control as soon as it approached a freight car, far enough in advance to enable the locomotive to slow down automatically to an impact speed of 1.2 mph, and then come back under tower control after impact or coupling.

4. To provide control from other locations, such as the receiving yard. The areas under control of each tower should be limited and defined by a "barrier" beyond which the control is automatically passed on to the next tower. If no signal is sent at that time, the locomotive simply stops.

5. To control the locomotive, when required, from any part of the yard by portable transmitters.

6. To enable normal manual control, to take over, should remote control fail, or should the locomotive be used outside the yard limits.

Four of these objectives have been attained and perfected:

1. The continuous control of speed over the entire range.

2. The control of forward and reverse movements.

3. The control of an approach and coupling speed.

4. Instantaneous changeover from automatic to manual operation.

Theory of the Automation Applied to Yard Locomotives

The locomotive chosen for the initial experiments was a diesel-electric unit of a type commonly used in the French Railway yards: a 6-axle Baldwin switcher weighing 110 tons and having a nominal rating of 660 horsepower.

Continuous Control of Speed

As on all diesel-electric locomotive, the regulation of the power plant is such that for a given speed of the diesel engine, the power delivered to the main generator has a definite value independent of the speed of the locomotive.

The power transmitted to the wheels is under the control of the diesel engine governor, actuated by a servomotor, which is controlled by the throttle. Varying the air pressure in this valve from 0 to 57 psi (pounds per square inch), by means of the throttle, enables the engineman to control the speed of the diesel engine, and thus the tractive effort of the locomotive.

Although the manual control of the diesel engine is, as a rule, quite flexible, two positions or steps in the control presented difficulties in the solution of this part of the problem. These were the cutting out of the resistance in the exciter field, and the shunting of the motor fields. These difficulties could have been eliminated by modifying the electrical transmission of the locomotive, but in order to retain the same manual operating conditions as for other locomotives of this same class, and also to prove that these difficulties could be overcome, these electrical transmission features were left unchanged.

The brake equipment consisted of an automatic pneumatic train brake, plus an independent pneumatic locomotive brake; the latter offered braking possibilities from 0 to 50 psi. The similarity between accelerating and braking controls provided the basic idea for the design of the speed-control mechanism. Both controls are operated by a single air pressure, referred to as pilot pressure, which is related to a single electric current of low intensity.

The remote-control equipment on the locomotive has three major components as described in the following.

1. The electropneumatic servomechanism which controls the diesel engine speed and the force of application of the brakes. This servomechanism is designed to function, when energized, in place of the existing manual controls, with no modifications of the conventional equipment of the locomotive. Thus, manual control is

Fig. 3. Variation of pressure p with respect to current i

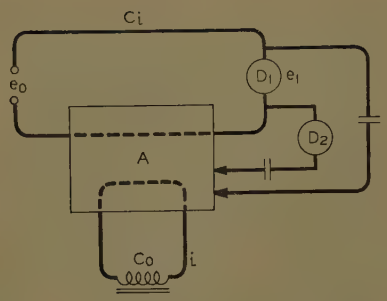


Fig. 4. Closed-loop circuit (simplified diagram showing how operating current i is produced)

e_0—Reference voltage related to required speed S_0

e_1—Voltage proportional to actual speed S of the locomotive, from D_1

D_1—Dynamo connected to the wheels of the locomotive $\left(\text{measures } L\frac{ds}{dt}\right)$

D_2—Dynamo measuring the speed (rpm) of the diesel engine $\left(\text{measures } M\frac{dw}{dt}\right)$

C_i—Input circuit to amplifier A

C_0—Output circuit of amplifier A

A—Amplifier

i—Operating current proportional to (S_0-S)

always available, and automatic control is made available instantly by simply turning a switch to energize it.

The mechanism consists mainly of two automatic feed valves, one for the engine, taking the place of the throttle, and the other for braking, taking the place of the independent brake. The working principles of the equipment require that the action of both of these valves be under the control of a single parameter, p, represented by a compressed-air pilot pressure, ranging from 0 to 57 psi. The higher values of this pressure activate the diesel engine control valve, and the lower values activate the brake valve.

The pilot pressure p is under control of an electric current i ranging in intensity from 0 to 150 milliamperes, through a small pilot feed valve operated by an electromagnet; see Fig. 2. The design of this valve is such that the pilot pressure p is always proportional to the intensity of the current i over the entire range of operations. In short, both engine and brake controls are subordinated to a single factor, the current i.

The use of compressed air, pilot pressure, for the actuating medium has proved to be especially flexible, and offers definite advantages for such an application. Compressed air is always available on a locomotive and affords the possibility of designing simple and rugged equipment which gives little trouble even though minor leakages may occur.

A reference to Fig. 3 will show more clearly the variation of the pressure p with respect to the intensity of the current i, and how these values are related to the engine and brake controls.

2. The electric speed-control mechanism, which relates the required speed S_0 of the locomotive to a voltage e_0, applied to the electric circuit of the servomechanism. This mechanism gives continuous control between 0 and 25 mph. Its working principle is based on a closed-loop circuit, see Fig. 4.

The reference voltage e_0, related to the required speed S_0, is opposed to another voltage e_i, proportional to the actual speed S of the locomotive, in the closed-loop circuit. The voltage e_i is supplied from a small dynamo connected to the locomotive wheels. The resulting electromotive force in this circuit, (e_0-e_i) is amplified by an amplifier A, which finally supplies in the output circuit C_0, a current i proportional to (S_0-S), the difference between the required and the actual speed.

Thus an equation is obtained:

$$i=K(S_0-S)+i_0$$

where i_0 and K are constants.

Refer now to Fig. 3 which shows the variation of engine or brake control pressure with respect to current i, and place i_0 in the above formula at the location i_0 on the graph. It will then be seen that when the actual speed S is smaller than the required speed S_0, current i is greater than i_0, thus causing the engine to accelerate, and the brakes to be released. When the actual speed S is greater than the required speed S_0, current i is smaller than i_0, thus causing application of the brakes; the engine will remain idle.

When S and S_0 are nearly equal, the currents i and i_0 are very nearly of the same value. Thus, neither engine nor brake is brought into action. Furthermore, by transposing i_0 in the above equation:

$$i-i_0=K(S_0-S)$$

where K is a constant, it is seen that both braking (when $i-i_0<0$) and accelerating (when $i-i_0>0$) are, as a rule proportional to the difference between the required speed S_0 and the actual speed S of the locomotive, at a given time. Thus, this mechanism provides a progressive action.

However, it must be noted that a mechanism of such simple design, intended for speed control, will lack stability, and is liable to set up sustained oscillations of speed around S_0. To avoid this phenomenon, the final equation should be written as follows:

$$i=i_0-K(S_0-S)-L\frac{dS}{dt}-M\frac{dw}{dt}$$

where $L(dS/dt)$ is a value proportional t the acceleration of the locomotive, bei derived from the small generator D_1, a where $M(dw/dt)$ is a value proportional t the angular acceleration of the diesel engi derived from a small generator D_2 on t engine.

The corrective term $M(dw/dt)$ necessary for two reasons:

(*a*). To avoid transitory racing of t diesel engine.

(*b*). To overcome the difficulties alrea mentioned of cutting out resistance in t exciter field and the shunting of the mot fields.

The degree of accuracy of the requir speed S_0 has certain limiting factors. On steep downgrade, with a heavy load, hol ing the speed S_0 will require high braki power. Hence, the current i will have very low value. On the contrary, on upgrade maintaining the speed S_0 w necessitate the full output of the die engine; and the current i will be arou the maximum value.

The difference ΔS between the act speed S on each of these two extre operating possibilities, and the requir speed S_0, will correspond to the variati of voltage Δe_0, which sets up through t amplifier A a variation of the current from minimum to maximum value.

This implies, theoretically, that t difference ΔS can be reduced as desir by a suitable increase in the amplificati factor. In practice, unstable conditio are liable to occur at given speeds wh the amplification factor is too high, th establishing certain limits. With t mechanism developed so far, ΔS is a proximately ± 1 mph, but further refi ments may be expected.

It is evident that the electromagne servomechanism, together with the el tric speed-control mechanism, perm speed control of the locomotive. T could be done manually on the locomoti by applying the required voltage e_0 to t circuit C_0 by means of a potentiometer, some similar apparatus. This ope teresting possibilities, since, witho

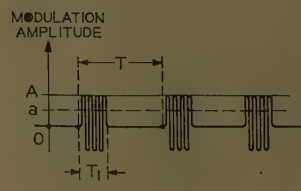


Fig. 5. Principle of interrupted modulati defining the remote-controlled speed as r=

ually going to full remote-control op-tions, the engineer's duty can be ilitated on humping operations, and a re accurate control of the speed can be tained.

3. The ratio transmission of the refer-e voltage e_0 from the control tower to locomotive. It has already been wn that an accurate and continuous ed control of the locomotive can be tained if such a voltage, proportional to required speed S_0, is applied to the vomechanism on the locomotive.

In the tower, a radio transmitter which erates a carrier wave of 160 megacycles nstalled. This carrier wave is modu-ed at 1,000 cps (cycles per second). e modulation is interrupted with a quency of some 15 cps so that during e of these interruption periods, T, the dulation frequency is maintained for a rtion of this time, T_1; see Fig. 5.

The ratio, $r = T_1/T$ gives the ratio of the note controlled speed S_0 to the maxi-m speed S_M of the locomotive:

$$\frac{T_1}{T} = \frac{S_0}{S_M}$$

The methods of producing such an errupted modulated carrier wave are ll known and required no details here. e advantage offered by such a prin-le is that the required speed is defined means of a number r, the ratio of two gths of time, T_1 and T, rather than by a ysical magnitude of the modulated rier wave, dependent on the quality of radio transmission. Furthermore, en radio transmission is not interrup-l, other modulated frequency bands of carrier may be used for transmitting ditional information, such as are needed the control of forward and reverse ections.

On the locomotive, the receiver is tuned that no current is set up during the ervals $T - T_1$, when no modulation is eived. During the interval T_1, when dulation is received, the receiver circuit illates with a constant amplitude in chronism with the transmitted modu-ion. This receiving device eliminates, erefore, any possible fluctuations in the nsmission.

In short, the reference voltage e_0 pro-ced in the receiver is proportional to the ration T_1 of the wave trains. By vary-this duration T_1 at the transmitter, voltage e_0 in the receiver is varied in same manner, thus controlling the ed S_0 as previously explained. There-re, it is apparent that the general erating principles of this speed-control uipment can be used on diesel locomo-es of various types.

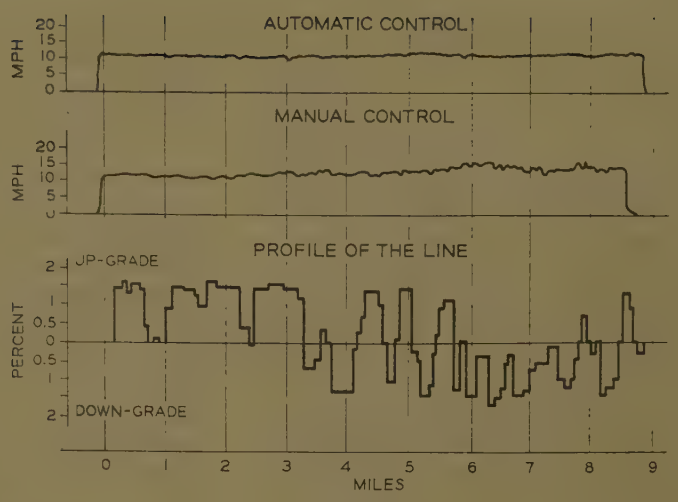


Fig. 6. Comparison between speed maintained by remote-control and by an experienced engineer

Control of Forward and Reverse Directions

The utilization of remote control on the directional equipment, forward and reverse motion, is achieved by using the same carrier as for speed control. Three low-frequency modulations are used:

110 cps for forward motion
125 cps for neutral position
140 cps for reverse

A remarkable feature of this equipment is that reversal of the direction can be ordered while the locomotive is in motion, without taking any special precaution. Memory relays, coupled with appropriate safety devices, allow the possibility of performing the operation automatically in the correct sequence.

Control of an Approach and Coupling Speed

The first solution of this problem which comes to mind is radar, which has not been used because it is too expensive and, furthermore, presents some difficulties in view of the short distances involved. Actually, detection must be made over a distance of about 100 feet in order to enable the locomotive running at 9 mph to slow down to 1.3 mph.

In a first test, the ultrasonic technique was used. The locomotive emits brief ultrasonic signals periodically which, after being reflected by the obstacle, are received by a microphone installed on the locomotive. The time elapsed between emissions and reception indicates the distance between locomotive and obstacle. Although the principle was satisfactory, there were irregularities in the performance due to parasitic echos and wind effects. Consequently, another method had to be sought.

The latest system developed electrically measures the length of track between the front axle of the locomotive and the rear axle of the car which is being approached.

A constant-frequency current, the intensity of which is inversely proportional to the length of the track, is induced in the track circuit. The current thus produces variable voltages in receivers mounted on the locomotive and, at a specific value of this voltage, two simultaneous actions are initiated:

1. The locomotive is momentarily cut off from the tower control.

2. The speed control is set automatically to an approaching speed of approximately 1.3 mph.

As soon as coupling is completed, the locomotive is brought back under radio control.

The locomotive is fitted with the following equipment:

1. For transmission, a loop is installed in a horizontal plane, 6 inches above rail level, in front of the locomotive. This loop, fed by a transistorized transmitter at a 50-kc frequency, is tuned electrically upon this frequency, with an output value stabilized at 100 megawatts, regardless of any variations of the supply voltage.

2. For reception, two receivers are installed in the same horizontal plane as the transmission loop. They consist mainly of ferrite-core coils, the purpose of which is to receive the magnetic component of the current induced in the track circuit. The position of these receivers is such that induction can be made only through the rails, and not directly from the transmitter loop.

The transistorized receiver equipment is composed mainly of a 2-step amplifier and a push-pull circuit which includes the control relay for the approach and coupling speed.

The present experimental application on the French National Railroads allows for an approach and coupling speed of 1.3 mph when the engine is about 100 feet from the car. Track curvatures slightly reduce this distance.

Achievements to Date

The practical application of the continuous control of speed was demonstrated experimentally with a test run on a main line where steep up and down grades were encountered; see Fig. 6. The locomotive hauled a dynamometer car equipped with radio transmitter, speed-control apparatus, and various recording instruments. A speed of 13 mph was maintained over the whole run with variations not exceeding ±1.2 mph.

The locomotive performed extremely well, fulfilling all of the expectations involved with the theoretical concepts. In fact, over the difficult route chosen, remote control performance was even better than manual control with an experienced engineman; Fig. 6.

The locomotive was then put in shunting service in a yard handling an average of 2,000 cars a day. At first, it was feared that unfavorable conditions might create difficulties for radio transmission. There is a forest in the immediate vicinity of the yard which creates a zone of absorption likely to produce interference and a concrete bridge crosses the incoming tracks where the locomotive must operate when a long train is being cut. Nevertheless, the locomotive executed all orders in an entirely satisfactory way.

To date, after 2 years of service, this locomotive is still in operation. Four additional remote controlled shunting units have been added since.

The French National Railroads is now working on another experimental application of remote control which consists of controlling the operating mechanism of a booster locomotive placed anywhere in a train from the head locomotive. In other words, the thought is to operate multiple units without any connecting wires between units. Since there is no limit to train loads because of coupler strength, this development will make it possible to run two or three trains together as a single unit. This might well be another step toward further development of automation in railroad operation.

Conclusions

The work carried out by the Research Department of the French National Railroads during the past 10 years has resulted in the introduction of remote controlled train operations as a current daily event in several classification yards. Even though this new development has not yet yielded the economic returns that might be rightfully expected, it is a step in the right direction toward the never-ending search for safety and improvement of operating efficiency.

Contact Wire Wear

KENNETH H. GORDON
MEMBER AIEE

CONTACT wire wear is important to the operator of an electric railroad, primarily because it contributes to the maintenance cost of the catenary system. The prudent operator will, therefore, periodically check the condition of the wire so that replacements may be planned in an orderly manner before breaks occur. The data presented here are based largely on results of such a survey, made on The Pennsylvania Railroad in1958, which covers approximately 94% of the electrified road track miles. Yard tracks were not included.

The entire catenary system operates at 11,000 volts, 25 cycles, single phase on this electrification. On so-called main-line tracks, a compound catenary system is used. This consists of a 5/8-inch bronze or composite-copper and copper-covered steel stranded messenger cable; a no. *4/0* Awg (American wire gage), grooved, hard-drawn copper auxiliary; and a single grooved, bronze contact wire. On most of the original construction, the size of this contact wire was no. *4/0* Awg. Later, to provide a longer-wearing life and because of some difficulty with fatigue breaks, larger wire was used for high-speed passenger tracks and on heavily travelled suburban tracks. A 300-MCM (thousand-circular-mil) cross section was used first but later, in order to increase the useful life further, the standard for these tracks was raised to a 336.4-MCM size, having an elongated, or somewhat oval, cross section. Size *4/0* Awg is still the standard for tracks normally assigned to freight trains. These three sizes of contact wire are illustrated in Fig. 1, which also indicates the amount of permitted wear and the remaining vertical diameter at which replacement is required.

On lines used exclusively for freight service (except for an occasional passenger train detour) a simple catenary system is used which consists of a 5/8-inch, stranded composite messenger cable, and a size *4/0* Awg, grooved, bronze contact wire. Lifting-type hangers are used to avoid hard spots. It is expected that the *4/0* Awg size will be continued when the time comes for replacement.

All-electric locomotives and multiple-unit cars collect current from the contact wire by means of pantographs equipped with mild steel shoes. No wearing strips are used. Road locomotive pantographs have two shoes and operate with an upward pressure of 32 pounds. Multiple-unit car pantographs have a single shoe and operate at a pressure of 22 pounds.

The principal cause of contact wire wear is probably simple mechanical abrasion from the pantograph shoe. Other probable contributing causes are amount of current collected, speed, and arcing, which may be the result of improper pantograph operation, some undesirable condition in the catenary construction, or contamination of the contact surface of the wire by dirt, soot, corrosion, or ice.

About the middle of 1940, a program of lubrication of pantograph shoes was inaugurated. This lubrication consisted of painting a coat of graphite on the shoe before installation and adding applications of the lubricant at terminals when time and opportunity were available. Results were measured in terms of pantograph shoe life. Some improvement was seen although not as great as had been hoped for. Intermediate application was difficult to control and the program has been largely discontinued except for the initial new shoe application. It was felt that the increased shoe life almost paid for the cost of the lubrication and that some saving from the probable decreased rate of wear of the contact wire resulted. The latter has never been specifically measured but it is believed to be beneficial.

Paper **61-190**, recommended by the AIEE La Transportation Committee and approved by t AIEE Technical Operations Department presentation at the AIEE Winter General Meeti New York, N. Y., January 29–February 3, 19 Manuscript submitted November 22, 1960; ma available for printing March 1, 1961.

KENNETH H. GORDON is with The Pennsylvan Railroad Company, Philadelphia, Pa.

The author wishes to express his appreciation C. H. Pagesy, J. H. Dean, S. V. Smith, J. Hogan, E. H. Brown, and L. B. Curtis of T Pennsylvania Railroad and to A. B. Costic The Delaware, Lackawanna and Western Railro for furnishing vital information.

The Illinois Central System started sing graphite lubrication late in 1935 in he form of a paste placed between two pper wearing strips on the pantograph noe. During the first 4 years, the life of earing strips was practically trebled and he contact wire wear (at a specific loca- on where measurements were taken nnually) decreased from a rate of 22.8 ils per 100,000 pantograph passes to bout 9.1 mils.

The Delaware, Lackawanna and Western Railroad Company started using graphite lubrication, similar to that of the Illinois Central, in 1937 and during the first 3 years, the life of the copper wearing strips increased approximately 50 per cent. A recent statement from the electrical engineer of that railroad indicates that the average life of wearing strips is now almost four times longer than before lubrication. Prior to lubrication, this company assumed that all size *4/0* Awg bronze, double contact wires would have to be replaced by 1950, after a life of approximately 20 years. During 30 years of operation, more than 131,000,000 pantograph miles have been accumulated. They have actually replaced 150,000 feet of the wire, or 14 miles out of a total of 157 miles of electrified track. This was largely due to improper initial installation methods, such as the use of improperly reeled wire, and thin spots at station platforms. An ultimate life of another 10 years, for a 40-year, total is expected for the rest of the wire. The vertical diameter at which the wire is condemned on this railroad is 250 mils. On this basis, the author estimates that the 40-year life represents an average wear rate of about 17 mils per 100,000 pantograph passes.

The Lackawanna, which has also been experimenting with carbon shoes, found that:

1. Carbon life matched that of the copper strips within 1,000 pantograph miles, the total being about 41,000 miles, and both ran for the same period.
2. There was no burning of the carbon.
3. While the test included two minor ice conditions, there was no evidence of any operating difficulty.
4. No special precautions or special maintenance were necessary.
5. Although carbons tended to groove and valley in the first few runs, they eventually smoothed out so that pantograph operation was not impaired.

If these shoes conform to present expectations, a further extension of the life of the contact wire may be realized.

On The Pennsylvania Railroad, the contact wire's rate of wear was not uniform throughout the electrified system but was most rapid near station platforms in suburban territory, where multiple-unit trains make frequent stops. This condition exists on most electric railroads and has been attributed to the high rate of current collection during acceleration of the trains. The Pennsylvania found that the rate of wear was about the same on both the approaching and leaving sides of the station, extending from 250 to 400 feet in each direction. However, many long trains are operated, so current collection may not be ruled out as a contributing factor. The rate of wear is 2 to 3 times as great in the stations as on the line between stops.

Approaches to low spots in the contact

Table I

Service	Approximate Wear Per 100,000 Pantograph Passes, Mils
Multiple-unit cars	10.2
Passenger locomotives	13.1
Freight locomotives	12.5

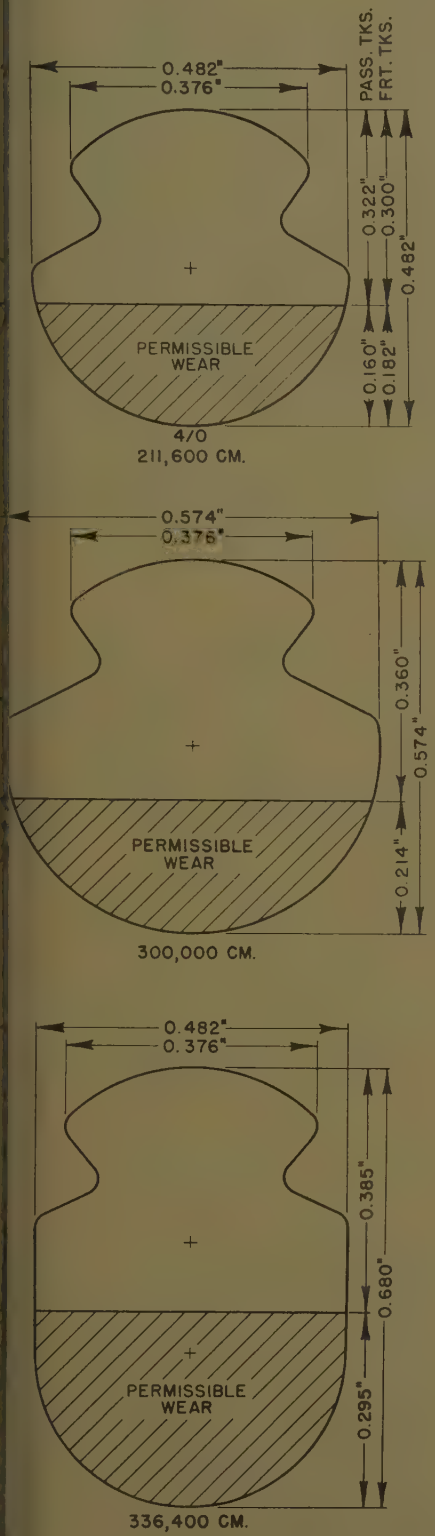


Fig. 1. Comparative size of contact wires

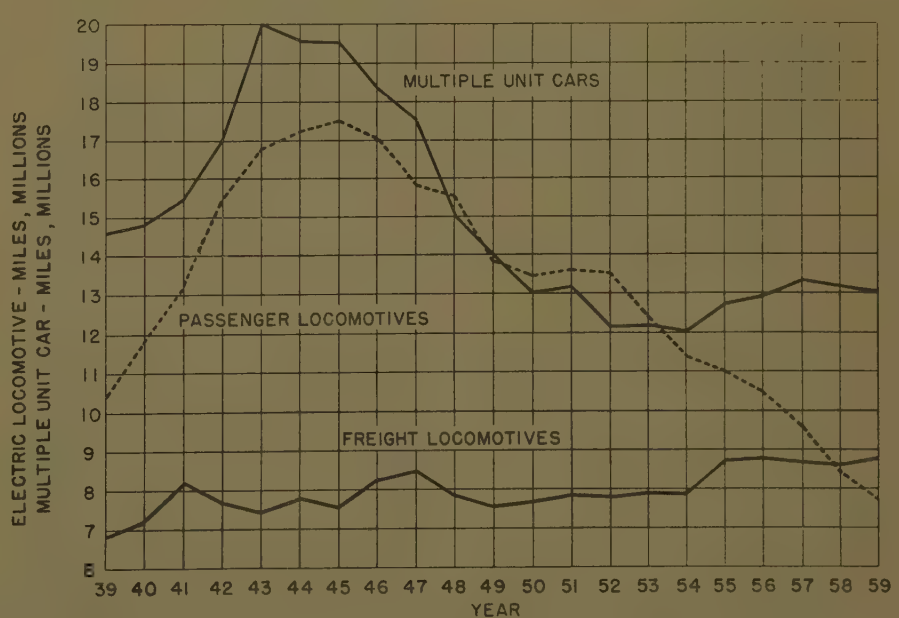


Fig. 2. Electric locomotive miles and multiple-unit car miles, 1939–59

Table II. The Pennsylvania Railroad Contact Wire Condition for 1958 and Planned Replacement, 1960 to 1984

Location	Track, Numbers	Date Installed	Size	Miles	Vertical Diameter, Mils	Annual Wear, Mils	Life, Years: Present	Life, Years: Remaining	Life, Years: Total	Planned Replacement, Miles	Passenger (P) or Freight (F) Tracks
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. New York region:											
1. East River tunnels	1 & 2	1932	300 MCM	5	450	4.8	26	19	45	5	P
2. East River tunnels	3 & 4	1932	300 MCM	5	505	2.7	26	54	80	none	P
3. North River tunnels	N & S	1932	300 MCM	5	480	3.6	26	33	59	none	P
4. Tunnel portals, Newark	1 & 2	1945	300 MCM	14	475	7.6	13	15	28	14	P
5. Newark–Rahway	1 & 4	1947	300 MCM	22	490	7.6	11	17	28	22	P
6. Newark–Rahway	2 & 3	1933	4/0 Awg	22	400	3.3	25	23	48	22	F
7. Rahway–Trenton	1 & 4	1947	300 MCM	74	530	4.0	11	43	54	none	P
8. Rahway–Trenton	2 & 3	1933	4/0 Awg	74	400	3.3	25	23	48	74	F
9. Trenton–Holmes	1 & 4	1953	336.4 MCM	38	650	6.0	5	44	51	none	P
10. Trenton–Holmes	2 & 3	1930	4/0 Awg	38	400	3.0	28	26	54	none	F
11. Jersey City Branch	1 & 2	1932	4/0 Awg	13	370	4.3	26	11	37	13	P
12. Other branch lines	All	1935	4/0 Awg	44	420	2.7	23	36	59	none	P & F
13. Other branch lines	All	1939	4/0 Awg	49	425	2.8	20	36	56	none	P & F
14. Total New York region				403						150	
B. Philadelphia region:											
1. Philadelphia–Paoli	1	1946	300 MCM	20	500	6.0	12	23	35	20	P
2. Philadelphia–Paoli	2	1933	4/0 Awg	20	420	2.5	25	40	65	none	F
3. Philadelphia–Paoli	3	1933	4/0 Awg	20	405	3.1	25	27	52	none	F
4. Philadelphia–Paoli	4	1946	300 MCM	20	495	6.6	12	21	33	20	P
5. Holmes–Philadelphia	1 & 4	1947	336.4 MCM	28	600	7.0	11	30	41	none	P
6. Holmes–Philadelphia	2 & 3	1930	4/0 Awg	28	405	2.8	28	29	57	none	F
7. Chestnut Hill Branch	1 & 2	1927	4/0 Awg	13	375	3.4	31	15	46	13	P
8. West Chester Branch	1 & 2	1949	300 MCM	37	535	4.3	9	40	49	none	P
9. Paoli–Harrisburg	All	1938	4/0 Awg	241	415	3.4	20	27	47	none	P & F
10. Columbia & A. & B. Branches and York Haven Line	All	1938	4/0 Awg	206	440	2.1	20	67	87	none	F
11. P. & T. and Trenton Branches	All	1938	4/0 Awg	113	440	2.1	20	67	87	none	F
12. Total Philadelphia region				746						53	
C. Chesapeake region:											
1. Brill–Bellevue	1 & 4	1947	300 MCM	44	515	5.4	11	29	40	none	P
2. Brill–Bellevue	2 & 3	1928	4/0 Awg	36	390	3.0	30	30	60	none	F
3. Bellevue–West Yard	2 & 3	1948	300 MCM	12	509	6.5	10	23	33	12	P
4. West Yard–Perryville	1 & 4	1934	4/0 Awg	32	430	2.2	24	49	73	none	F
5. West Yard–Perryville	2 & 3	1955	336.4 MCM	8	662	6.0	3	46	49	none	P
6. West Yard–Perryville	2 & 3	1954	300 MCM	8	550	6.0	4	31	35	none	P
7. West Yard–Perryville	2 & 3	1934	4/0 Awg	47	380	4.3	24	14	38	47	P
8. Perryville–Gunpow	1	1934	4/0 Awg	4	410	3.0	24	29	53	none	F
9. Perryville–Gunpow	2	1934	4/0 Awg	3	360	5.1	24	8	32	2	P
10. Perryville–Gunpow	2	1952	300 MCM	4	530	7.3	6	23	29	4	P
11. Perryville–Gunpow	2	1953	336.4 MCM	12	650	6.0	5	44	49	none	P
12. Perryville–Gunpow	3	1952	300 MCM	4	530	7.3	6	23	29	4	P
13. Perryville–Gunpow	3	1952	336.4 MCM	4	643	6.2	6	42	48	none	P
14. Perryville–Gunpow	3	1934	4/0 Awg	11	360	5.1	24	8	32	11	P
15. Perryville–Gunpow	4	1934	4/0 Awg	4	400	3.4	24	23	47	4	F
16. Perryville–Gunpow	4	1943	300 MCM	9	515	3.9	15	39	54	none	F
17. Gunpow–Baltimore	1	1944	300 MCM	6	518	4.0	14	39	53	none	F
18. Gunpow–Baltimore	1	1934	4/0 Awg	10	385	4.4	24	16	40	10	F
19. Gunpow–Baltimore	2 & 3	1953	336.4 MCM	1	650	6.4	5	44	49	none	P
20. Gunpow–Baltimore	2 & 3	1934	4/0 Awg	31	360	5.1	24	8	32	31	P
21. Gunpow–Baltimore	4	1943	4/0 Awg	7	406	5.1	15	16	31	7	F
22. Gunpow–Baltimore	4	1934	4/0 Awg	9	390	3.8	24	18	42	9	F
23. Baltimore–Landover	1	1947	300 MCM	1	497	7.0	11	19	30	1	P
24. Baltimore–Landover	3	1934	300 MCM	1	478	4.4	24	29	53	none	F
25. Baltimore–Landover	2 & 3	1956	336.4 MCM	1	668	6.0	2	47	49	none	P
26. Baltimore–Landover	2 & 3	1934	300 MCM	3	405	7.0	24	6	30	3	P
27. Baltimore–Landover	1	1934	4/0 Awg	25	400	3.4	24	23	47	25	F
28. Baltimore–Landover	4	1934	4/0 Awg	13	390	3.8	24	18	42	13	F
29. Baltimore–Landover	2 & 3	1956	300 MCM	15	562	6.0	2	33	35	none	P
30. Baltimore–Landover	2 & 3	1934	4/0 Awg	47	360	5.1	24	7	31	47	P
31. Landover–Washington	2 & 3	1934	4/0 Awg	12	398	3.5	24	22	46	12	P
32. C. & P. D. Branch	All	1937	4/0 Awg	52	440	2.0	21	70	91	none	F
33. Total Chesapeake region				476						242	
D. Total three regions				1,625						445	

wire, such as at overhead bridges, was another location where rapid wear was initially experienced. However, this has been brought under control by extending the length of the gradient approaching such low spots and thereby decreasing the steepness.

Aside from these locations, the rate of wear appears to vary with the number of pantograph passes and the type of service operated under the wire. Unfortunately, there is no readily available record, either in total or by class of service, of the number of pantograph passes occurring in any specific area in electrified territory. However, by utilizing the data which are available, an approximation may be made.

For this purpose, 1950 was selected as an average year during the life of the electrification. Official timetables were examined and the number of scheduled passenger trains, hauled by both locomotives and multiple units, was determined for each of several specific sections. From other records the average number of locomotive units per train (1.06) and of multiple-unit cars per train (4.93) were determined. The distribution of the locomotive and multiple-unit car miles and the resulting pantograph passes, were then estimated. The average number of freight locomotive miles, per mile

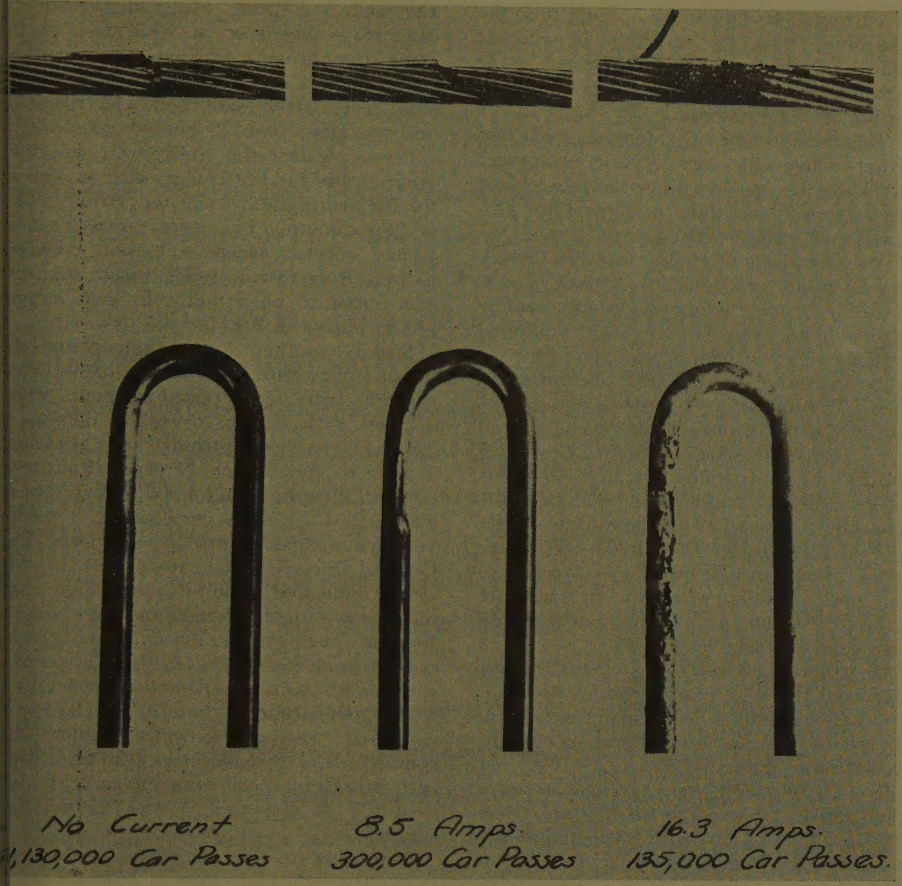


Fig. 3. Effect of excessive current feeding through loop hangers

reight track, was also determined from ne records and it was assumed that the novement over the principal freight ranches, which are used exclusively for reight service, was about equal to the verage. Since the original wire is still in lace over the freight branches, and since freight movement with electric locomotives has been reasonably uniform, the latter figures were satisfactory for use. However, much of the present contact wire in the selected passenger train areas was installed only a few years prior to 1950 and some of it since that year. Therefore, the average passenger train figures had to be modified to take into account the changing traffic pattern of recent years; this is shown in graph form in Fig. 2.

The estimated average pantograph passes per year were applied to the average contact wire wear per year, and the results are shown in Table I.

The derivation of these approximations required several estimations and assumptions and the values should, therefore, be used with caution. However, they are based on the best available information, and it is believed that they are satisfactory for normal catenary construction. They are not applicable, however, to the wire in the New York tunnels and possibly other special locations.

Data regarding the contact wire on 1,625 of the 1,723 electrified road-track miles, excluding 426 miles of electrified yard and siding track, are given in considerable detail in Table II. Installation date, size, length, vertical diameter, average annual wear, and life of the wire, as well as the planned replacement program for the 25-year period from 1960 to 1984, are included in the data. It will be noted that about 27% of the mileage is scheduled for replacement during this period, or a little over one % per year. About 63% of this is on passenger tracks, some of which may be postponed further, in view of the reduced level of passenger traffic.

The bulk of catenary maintenance cost is preventive maintenance, for inspection, minor repairs, etc., but major replacements of contact wire add substantially to this cost. The average annual cost of planned replacements over the 25-year period would amount to about 14% of the annual cost without major replacements. Both this cost and the inconvenience of such replacements can be held to a minimum if replacements are carefully planned and programmed over the years rather than permitting the wire to wear to the point where emergency measures become necessary.

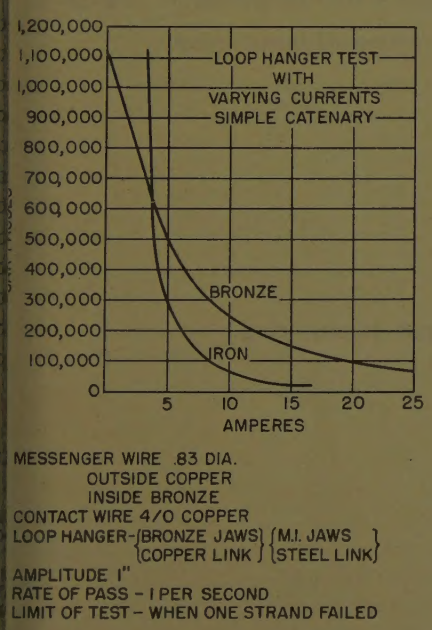


Fig. 4. Loop hanger test

Discussion

L. W. Birch (Ohio Brass Company, Mansfield, Ohio): Having been actively engaged with transit and railroad overhead distribution problems for many years, I can appreciate the great effort and accumulation of knowledge represented in Mr. Gordon's paper, which should be useful to many companies, both domestic and foreign, who are concerned with electrified lines. With respect to Mr. Gordon's

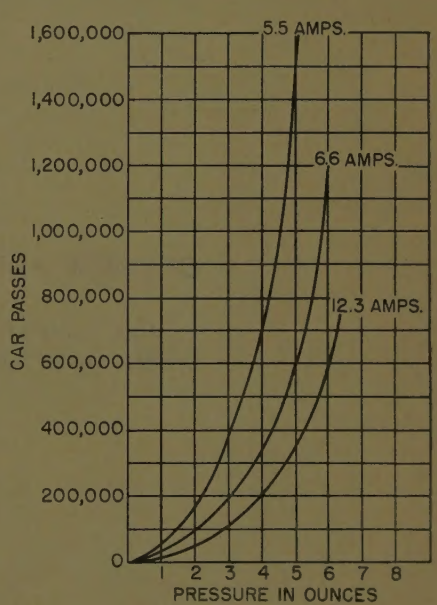


Fig. 5. Life of steel loop hanger with varying pressures against messenger

statement that the "principal cause of contact wire wear is probably simple mechanical abrasion from the pantograph shoe," I would like to add some comments, which may throw additional light on this subject.

Some years ago, we were confronted with loop catenary hanger burning on several railroads because of an excessive amount of current passing from the messenger cable, through the loop hanger, and into the trolley wire. The loop hanger is selected because it provides flexibility due to its ability to lift from the messenger cable. This sliding, lifting action can cause considerable damage because of the breaking of the electric circuit between the hanger loop and the messenger cable when a pantograph passes beneath the hanger.

A test, Fig. 3, was conducted and, on the basis of its results, we were able to vary the vertical amplitude of the hanger, as well as simulate the speed of the train. Currents ranging from 0 to 16 amperes were passed through the hanger, and various materials, both ferrous and nonferrous, were used in the experiment. Some interesting figures were obtained from the 5 years of continuous, 24-hour-per-day experiment.

At 15 amperes, Fig. 4, an iron loop hanger lasted 30,000 pantograph passes, while the bronze loop hanger lasted 150,000. At 10 amperes, the iron hanger was extended to 70,000 passes, while that of the bronze reached 250,000 passes. At 8 amperes, the respective passes were 110,000 and 320,000. At 3.5 amperes, both hangers withstood 600,000 pantograph passes, which is indicated by the point where the two life curves pass in the illustration. At zero amperes, when there was only mechanical abrasion to affect the hanger rods, the iron hanger loop withstood 13,000,000 pantograph passes while the bronze rod withstood only 1,100,000. Thus, a variation in the life of the hanger results from both abrasion and the passage of current. It should be noted, from the foregoing, that the amount of current which passed from the messenger cable into the hanger loop was relatively small during all experiments.

The termination of a test was reached either when one wire of the messenger cable was severed or when the hanger rod itself became too thin to support its proportion of trolley wire.

Another phase of trolley wire life history has been presented by the trolley coach. A no. *2/0* Awg grooved, bronze trolley wire, having the swivel harp equipped with a trolley wheel, which was used for trolley coaches until 1930, permitted a life of approximately 400,000 coach passes. Then, the old bronze wheel was replaced by a steel shoe which reduced the life of the *2/0* grooved trolley wire to 75,000 bus passes, without lubrication. By lubricating the trolley wire, the steel shoe's life was increased to 500,000 car passes. After 1935, the carbon shoe insert was introduced, and now a trolley wire's life in excess of 3,000,000 coach passes for *2/0* grooved wire can be expected.

On their new east–west rapid-transit line, the Cleveland Transit System installed a single metallic shoe pantograph to operate on a catenary system which included only one trolley wire. The old rule-of-thumb tells us that one contact point made by a single pantograph shoe and a single trolley wire will enable the collector to handle approximately 500 amperes, not 1,000 amperes which was normally collected. Extreme trolley wire wear at certain locations was experienced with this arrangement. Ultimately, the pantographs were replaced with a 2-pan collector equipped with carbon blocks. The previously experienced trolley wire wear, which was attributed to burning, was reduced to a minimum and the wire life was extended.

With respect to the pressures of a collector against the trolley wire, there seems to be a minimum pressure which will prevent the collector from being disturbed by high train speeds or by wind and, at the same time, provide adequate commutation. The Illinois Central Railroad made extensive tests after suburban electrification and decided that 21-pound pressure of collector against the trolley wire provided greatest life for the trolley wire as well as for the pantograph collector strips. This is approximately the net pressure used on trolley coaches where a carbon collector is pressed against a trolley wire. Pressure was found to be a factor in loop hanger life, as shown in Fig. 5.

Speed, another factor influencing the life of trolley wire, does not, in itself, cause greater wire deterioration, but when coupled with certain overhead defects it can cause severe burning and abrasion. I refer to the French National Railway's tests of 1954, in which 4,000 amperes were collected at 142 miles per hour on a 1,500 volt line without injury to the wire. The General Electric Company's test in the 1920's indicated good life with high currents, but both the French and the General Electric engineers realized that certain modifications had to be made to catenary systems to secure satisfactory collection. Today, smoothness, uniform inertia, and restricted vertical deflection must be considered for the overhead system, otherwise satisfactory collection of current and long life of the trolley wire cannot be attained.

Returning to Mr. Gordon's statement that "the principal cause of contact wire wear is probably simple mechanical abrasion from the pantograph shoe," I agree that electrical wear or burning can be kept to a minimum with adequate trolley wire tension, bumpless overhead, and controlled current collection. By the latter term, I mean within the limits of the collector for various values of current. Nevertheless, it should be remembered that current is responsible for a considerable amount of deterioration of grooved trolley wire.

Power Apparatus and Systems—August 1961

AIEE PUBLICATIONS

	Member Prices	Nonmember Prices: Basic Prices*†	Nonmember Prices: Extra Postage for Foreign Subscription
Electrical Engineering			
Official monthly publication containing articles of broad interest, technical papers, digests, and news sections: Institute Activities, Current Interest, New Products, Industrial Notes, and Trade Literature. Automatically sent to all members and enrolled students in consideration of payment of dues. (Members may not reduce the amount of their dues payment by reason of nonsubscription.) Additional subscriptions are available at the nonmember rates.		annually $12* Single copies $1.50*	$1.00
Bimonthly Publications			
Containing all officially approved technical papers collated with discussion (if any) in three broad fields of subject matter as follows:			
	annually	annually	
Communication and Electronics	$5.00	$8.00*	$0.75
Applications and Industry	$5.00	$8.00*	$0.75
Power Apparatus and Systems	$5.00	$8.00*	$0.75
Each member may subscribe to any one, two, or all three bimonthly publications at the rate of $5.00 each per year. A second subscription to any or all of the bimonthly publications may be obtained at the nonmember rate of $8.00 each per year.			
Single copies may be obtained when available.	$1.50 each	$1.50* each	
AIEE Transactions			
An annual volume in three parts containing all officially approved technical papers with discussions corresponding to six issues of the bimonthly publication of the same name bound in cloth with a stiff cover.			
	annually	annually	
Part I Communication and Electronics	$4.00	$8.00*	$0.75
Part II Applications and Industry	$4.00	$8.00*	$0.75
Part III Power Apparatus and Systems	$4.00	$8.00*	$0.75
Annual subscription to all three parts (beginning with vol. 77 for 1958).	$10.00	$20.00*	$2.25
Annual subscription to any two parts.		$15.00*	$1.50
AIEE Standards			
Listing of Standards, test codes, and reports with prices furnished on request.			
Special Publications			
Committee reports on special subjects, bibliographies, surveys, and papers and discussions of some specialized technical conferences, as announced in ELECTRICAL ENGINEERING.			

*Discount 25% of basic nonmember prices to college and public libraries. Publishers and subscription agencies 15% of basic nonmember prices. For available discounts on Standards and special publications, obtain price lists from Order Department at Headquarters.

†Foreign prices payable New York exchange

Send all orders to:

Order Department
American Institute of Electrical Engineers
345 East 47th Street, New York 17, N. Y.